Report No. 4

FINAL REPORT

1 Julv 1960 to 30 September 1961

SPEECH BANDWIDTH COMPRESSION SYSTEM

(Scanning Filter)

Contract No. DA-36-039-sc-85-140

DA Project Nr. 3B-31-01-001

Prepared for U. S. Army Signal
Research and Development Laboratory
Fort Monmouth, N. J.

Object of Research: To perform an engineering study of speech bandwidth compression by using filter scanning techniques.

Signal Corps Technical Requirements

Nr. SCL 4214, 17 February 1960
Nr. SCL 2101K, 15 July 1959

Contributors:

A. P. Albanese
N. Wand
B. Zivatofsky
L. Zolotnitsky

PRD Report 341.0911 FR
September 1961

PRD ELECTRONICS, INC.
A Subsidiary of Harris-Intertype Corporation
Brooklyn 1, New York

CONTENTS

ILLUSTRATIONS

ILLUSTRATIONS (Cont'd)

TABLES

1. PURPOSE

The purpose of this project is to conduct an engineering study of Speech Bandwidth Compression Systems using the method of a frequency scanning filter for sampling the speech frequency-time spectrum and using reverberation techniques for interpolation between samplings.

2. ABSTRACT

This report describes the analysis, experimentation and fabricational activity directed toward a time-frequency scanning speech compression system. A statistical analysis was made to determine the scanning waveforms based on articulation scores for selected zero crossings of infinitely clipped speech. In addition, an empirical approach using sawtooth, triangular, sine-wave and rectified sine-wave scans was employed to determine the sampling scan and rates which would yield maximum articulation. Experimentation shows that using an Autovox for reverberation improved articulation. This project initiated the concept of a speech compression system using two filter scans to intercept formant changes and yield maximum articulation.

3. PUBLICATIONS, REPORTS, AND CONFERENCES

3.1 PUBLICATIONS - None.

3.2 LECTURES - None.

3.3 REPORTS - None.

3.4 CONFERENCES

3.4.1 The first project conference was held on 21 July 1960 at PRD between Mr. Joseph De Clerk of your laboratory and Mr. Angelo P. Albanese of this laboratory. The purpose of this meeting was to discuss policy of the contract and to review some of the technical aspects of the project.

3.4.2 A project conference was held on 7 September 1960 at our plant between Mr. Martin Weinstock and Mr. Joseph De Clerk of your laboratory and Mr. Angelo P. Albanese of this laboratory. This meeting outlined some of the areas to be investigated on this project.

3.4.3 A conference was held on 3 October 1960 to discuss Speech Bandwidth Compression Systems. The personnel present were: Mr. Martin Weinstock of your laboratory, Dr. Raisbeck* of Bell Telephone Laboratories, Mr. Kaiser of IDA, Mr. Solee* of NSA, Dr. Carlos Angulo* of Brown University, and Dr. M. J. DiToro and Mr. Angelo P. Albanese of PRD. The purpose of this meeting was to determine the present status of work performed on Speech Compression Contracts. The visiting group was concerned with the progress of speech compression, possibility of improvements, and the overall program in the United States which is devoted to speech compression systems. The formal discussion was mainly concerned with the technical proposal written by Dr. M. J. DiToro.

---

* On loan to the Institute of Defense Analysis.

3.4.4 A project conference was held on 9 December 1960 at our plant between Mr. Joseph De Clerk of your laboratory and Mr. Angelo P. Albanese of this laboratory. This meeting supplied Mr. De Clerk with information of progress to date, and an outline of future areas of investigation.

3.4.5 A project conference was held at our plant on 1 February 1961 between Mr. Joseph De Clerk and Mr. Frederick Evans of your laboratory, and Mr. Angelo P. Albanese of this laboratory. The purpose of this meeting was to discuss the project's future status without the consulting services of Dr. M. J. DiToro. Mr. De Clerk had been assured that the requirements as stated in the proposal would be delivered by PRD.

3.4.6 Another project conference was held on 9 March 1961 at our plant between Mr. Frederick Evans of your laboratory, and Mr. Angelo P. Albanese of this laboratory. This meeting supplied Mr. Evans with information on the progress of the project and its future areas of investigation.

3.4.7 A project conference was held on 4 May 1961 at our plant between Mr. Fred Evans of your laboratory and Messrs. Angelo P. Albanese, Leon Zolotnitsky and Bernard Zivatofsky of this laboratory. Individual conferences were held to display the equipment and report on project progress to date. Future endeavors were outlined and discussed with anticipated system results.

3.4.8 A project conference was held on 25 May 1961 at our plant between Messrs Martin Weinstock, Joseph De Clerk, and Fred Evans of your laboratory and Dr. L. S. Castriota and Angelo P. Albanese of this laboratory. This conference was a demonstration of the speech compression system and a discussion on future modification of the present equipment. Another topic

of discussion was a presentation of a new concept for achieving speech compression. Details were disclosed concerning a new and improved speech compression system using two scanning filters to sample the speech spectrum. At this time, PRD had asked for an extension of three months to complete work on the project at no extra cost to Fort Monmouth.

3.4.9 A project conference was held at our plant on 17 July 1961 between Mr. Fred Evans of your laboratory and Messrs Angelo P. Albanese and Leon Zolotnitsky of this laboratory. This meeting informed Mr. Evans of the progress to date and related the extension program.

3.4.10 A project conference was held at our plant on 27 September 1961 between Mr. Joseph De Clerk of your laboratory and Dr. L. J. Castriota and Mr. Angelo P. Albanese of this laboratory. The final developments and results of this project were discussed. Delivery of the project system with technical procedures were arranged.

4. FACTUAL DATA

4.1 INTRODUCTION

The application of speech bandwidth compression to voice communication promises better use and improved performance of existing communication links. With increased traffic demands on these communication links, channel capacity economy is desirable.

Speech bandwidth or channel capacity reduction systems are normally divided into four groups*. (1) Time or frequency compression: Sampling or frequency division techniques with a bandwidth compression of four permit a binary channel capacity of 5,000 to 10,000 bits per second. This project falls within this category. (2) Continuous analysis-synthesis: Transmission of analog control signals in place of speech signals yields bandwidth compression of about 2,000 bits per second. (3) Discrete sound analysis: Here speech signal code groups substitute for the speech signal and identify fundamental sounds eliminating emotional and personal cues. System capabilities involve information rates as low as 60 bits per second. (4) Sound group analysis-synthesis: Transmission of certain words or phrases identified by code groups have estimated useful information rates of 5 to 10 bits per second. However, such rates are a function of the scope of vocabulary used.

4.2 THEORY OF FILTER SCANNING

The speech compression technique of this project is best shown by referring to figure 1. Here an energy vs frequency vs time spectrograph of a voiced speech sound is displayed. The pattern shown in the spectrograph is sampled by the scanning filter of bandwidth Δf. Time per scan is designated as T and the range of speech frequency scanned is 250 to 3250 cps.

---

* S. J. Campanella, A Survey of Speech Bandwidth Compression Techniques, IRE Trans Audio, Sept - Oct 1958, p. 105

Figure 1b is a representation of filter scanning. Figure 1c shows the output energy of this filter, indicating that it contains sufficient sampling to permit an approximate reconstruction of the original spectrograph. Figure 1d shows the restored spectrum, where with reverberation, interpolation between samples is achieved.

A block diagram of a basic system performing these functions is indicated in figure 2. At the transmitter the speech spectrum input of 250 to 3250 cps is scanned, and the compressed speech signal shifted to a common center frequency f_a or 20 kc. This upward translation in frequency is effected with a balanced modulator and a sawtooth scan of the local oscillator. The sawtooth scan and the resultant spectrum shift is shown in figure 3a. The sawtooth scan illustrated is merely representative since other waveforms may be generated for scanning by the function generator.

The scanning rate $\frac{1}{T}$ of the function generator is controlled by a fixed sinusoidal synchronizing signal of frequency f_s derivable from $\frac{1}{T}$. As stated previously, the compressed speech signal has shifted to a common center frequency f_a or 20 kc. The waveform for this compressed signal is shown in figure 3c, while figure 3b shows the original speech signal. The fixed synchronizing frequency f_s is chosen to be at either edge of the link bandwidth and so can be readily transmitted and consequently extracted.

At the receiver the compressed signal is subjected to an inverse process. The sampled spectrum is approximately restored to the original speech spectrum by means of reverberation. The reverberators used are commercial units such as Fisher's Spacexpander and Kay's Autovox, and a feedback amplifier.

Restoration is possible if the sampling interval T of Figure 1b is the reciprocal to twice the bandwidth* Δf_c of the signal vs time representing the formant center frequency. The compressed bandwidth ratio for this system is $\sqrt{f_o/(2f_c)}$, where f_o is the original bandwidth of the speech signal and f_c is the effective cutoff

---

*C. E. Shannon, Communication in the Presence of Noise, Proc. IRE, Jan. 1949, p. 10.

frequency vs time signal. With restoration by means of reverberation, this system improves articulation by providing a continuity-in-sampling to the ear-brain chain.

4.3 SYSTEM ANALYSIS

The major difference between the general and specific project system (figures 2 and 4) is that the latter system employs a single scanning oscillator for purposes of system simplicity. This project envisioned no physical separation of transmitter and receiver sections, so that the basis of compression could be illustrated with the use of a single scanning oscillator. Explanatory system details follow.

First, the Krohn-Hite filter limits the speech spectrum input from 250 to 3250 cps. This serves to transmit the first three formants which contain information yielding articulation scores of better than 95%. A reference bandwidth of 3 kc will serve to determine compression ratios.

Simultaneously, the function generator and scanning oscillator produce a frequency-modulated signal. The function generator furnishes a variety of scanning waveforms which modulate the local oscillator, the scanning oscillator being an astable multivibrator whose base return voltage is changed by the output of the function generator. The asymptotic voltage created by this scanning waveform will generate a variety of frequencies. Although the oscillator waveform at the collector is a square wave, the balanced modulator and filter combination will only permit the fundamental frequency to mix with the speech spectrum.

Referring now to the schematic in figure 5, the scanning oscillator is coupled to the balanced modulator via transistor and transformer combination. Besides providing a proper amplitude level for modulation, the balanced modulator serves as an impedance match. Since the balanced modulator is composed of matched diodes in a bridge circuit, the diodes are switched on and off by the polarity of the incoming signals resulting in carrier suppresion.

The narrow bandpass filter is an L-C network that samples the speech spectrum in a manner dictated by the scanning waveform. Normally, amplitude response is the criterion of filter performance, but here phase response or phase deviation from linearity can contribute to frequency dispersion and impair the signal. Consequently, the signal output of the filter and signal output of the scanning oscillator would not be time synchronized. The delay within the narrow bandpass filter would, therefore, produce a signal at the receiver portion uncorrelated to the transmitter output. The Golay delay line, an adjustable L-C network compensates the average phase variation, and assures correct restoration of the speech spectrum at the output of the demodulator. Spectrograph figure 6 illustrates the distortion without the use of the delay line and its correction when the delay line is added.

Obviously, at the output of the demodulator, the audio spectrum appears with the addition of the higher frequency sidebands. These sidebands are attenuated by the Krohn-Hite bandpass filter. Although, the signal appearing at the output of this filter is continuous in time, spectral time samples develop as a result of the scanning rate and the narrow bandpass filters.

Referring now to figure 4, after the speech spectrum signal has been restored by demodulation, the signal enters the reverberator. Here the Autovox reiterates the sampled spectrum one or more times. The number of repetitions will be determined by the compression ratio, waveform, etc. The amplifier paralleling the Autovox ex-

tends the number of repetitions to three. Ballantine amplifiers appearing throughout figure 4 amplify and match impedances at locations shown.

4.3.1 Scanning Waveforms.

Before we can choose the scanning waveform which will sample the speech spectrum, we must know something about the information content of speech. Obviously, no information can be extracted if we observe the amplitude versus time waveform of speech, other than the fundamental pitch. However, by observing the sound spectrograph we get a representative display of where the information exists. If we observe a typical sound spectrograph figure 1a, it can be seen that the formants present will be time variable. The fact that each formant changes frequency constitutes modulation of sound contributing to intelligibility. Variations of sound are attributed to the change in frequency of the high intensity formants. Although, we cannot predict where these changes occur we can attempt to choose a scanning waveform based on previous experimentation.

If there is absolutely no correlation of the scanning waveform to the changes of the formant bars, then linear scanning by sawtooth waveform will serve to sample information content with the same degree of effectiveness as any other waveform. However, since this is uncertain, we can only attempt to use other waveforms and compare the results experimentally.

Another aspect involves the discontinuity present in the sawtooth scan which will produce high order transients resulting in low articulation scores*. To overcome this and still maintain an equal time sample, the triangular scan can be used to reduce these undesirable transients. However, since the lower formant contains the highest contribution to intelligibility, it would seem advantageous to use a waveform

*D. L. Subrahmanyam and G. E. Petersen, Time-Frequency Scanning in Narrow-Band Speech Transmission, IRE Trans. Audio, Nov - Dec 1959, p. 148.

that would scan the lower formant for a longer period of time. Such a typical waveform is the negative half-wave rectified sine wave. A mathematical treatment for this waveform is contained in Appendix B of the First Quarterly Report. Again, since no a priori information is available as to information content of speech, the positive half-wave rectified sine-wave scan is used for comparative purposes. Even these rectified scans present non-linear sampling with some discontinuity*. Therefore, the sinusoidal waveform with its inherent smooth transitions will be used to emphasize maximum sampling time at the lower and higher formants.

Depending on the scan used, the reverberator must fill in the gap between samples. In the case of the linear type of scan, e.g., sawtooth, triangular, -- with reverberation, the entire spectrum will be continuous. However, in non-linear scanning, the reverberator will fill in the gap, will not fill in the gap and/or will overlap the gap between samples. No mathematical analysis will be evaluated for the reverberator; however, this factor will be governed and determined experimentally.

As described, the waveforms discussed contain no known correlation to the information content of the speech spectrum. In order to develop some correlation, the following statistical method of inquiry was used.

4.3.1.1 Statistical Analysis of Speech Spectrum.

Experimental investigation shows differentiated and infinitely clipped speech has an articulation score of 95%**. From this we would infer that the information lies in the relative positions of successive zero crossing of differentiated and infintely clipped speech signals. Application of the spectral content of these zero crossings will be used to generate a scanning waveform which will "on the average" sample the speech spectrum optimally. It is therefore necessary to resolve the

---

* Here discontinuity is defined as the point which has no derivative as is apparent at the cusp.

** J. Licklider and I. Pollack, Effect on Differentiation, Integration and Infinite Clipping Upon Intelligibility of Speech. JASA, 1948, vol. 20, p. 42.

problem statistically. Since the zero crossings are correlated to the maxima and minima of speech amplitude, statistical results can be achieved simply by determining the relative frequency of occurrence for specified periods. The ultimate issue can therefore be framed in the question:"What is the probability density of having a zero crossing with positive slope at time t_1 if it is known that there exists a zero crossing with negative slope where t is equal to 0 ?"

To establish this probability density of time lapse between successive events, i.e., time of successive zero crossings, the experimental apparatus and schematics, figures 7, 8 and 9, are used. Essential circuits include the clipper amplifier and distribution analyzer.

An explanation of this circuitry may be facilitated by referring to these figures. With the exception of the clipper amplifier, all the circuitry is transistorized. In the clipper amplifier, each stage is preceded by a pair of silicon diodes which limit the voltage excursion at the tube input. The circuitry of the distribution analyzer, includes the delay multivibrator, triggering circuits for the pulse generator, and an "and" gate. With this circuitry, the differentiated speech signal is clipped and amplified and converted to a random series of square waves. Differentiation of the square waves produces positive and negative spikes which are time markers for the location of positive and negative slope zero crossings of the processed speech signal. The negative pulse is delayed for a period of t_1 by the delay multivibrator whose trailing pulse output triggers pulse generator no. 1. A pulse width Δt is produced which activates the "and" gate.

The succeeding positive spike at the output of the clipper amplifier triggers pulse generator no. 2 which produces a narrow pulse. If the time lapse between these two spike waveforms is between t_1, and $t_1 + \Delta t$, the output from pulse generator no. 2 will find the "and" gate open and will be registered by the counter. If the positive spike follows the negative spike by less than time t_1, this latter pulse will reset the multi-

vibrator and prepare the system for the next pair of pulses. In this way, no zeros are lost and a reasonably accurate distribution can be obtained. The period t_1 is set by a variable capacitor in the multivibrator and for a particular setting establishing t_1, the pulse count will be N_1.

Thus, if the total number of pairs of zeros in the speech sample is N_T, the probability of the distance between adjacent zero between t_1, and t_1 and Δt is $N_1/N_T \Delta t$. By taking data for different values of t_1, the entire density function may be obtained (figure 10).

The probability density curve for maxima-minima spacings is shown in figure 10a. If therefore, clipped differentiated speech passes through a pulse width selector set for a pulse width between t_2 and t_1, only those pulses whose width lies in this range will pass. (See figure 10b.) Then, the average rate of pulses at the output will be the product of the shaded area of the curve and the average pulse rate in the original signal. This involves dividing the curve into a number of equal areas as indicated and obtaining articulation scores for each section by means of the pulse width filter. The equipment can be modified slightly to function as a pulse width selector since the counter need only be replaced by a monostable multivibrator.

Pulse generator No. 1 is adjusted for a pulse width t_2 and t_1. For every pulse in the clipped speech with a pulse width between t_2 and t_1, the multivibrator will be triggered. It will put out a pulse of width:

$$\frac{1}{t_2 - t_1} \int_{t_2}^{t_1} t\, p(t) dt$$

This is the average pulse within the interval*.

---

*This integral is the "expected value" of the pulse width for pulses whose widths are within the specified range.

The output waveform will be a pulse-code-modulated signal. This signal will be used to measure the articulation of the areas shown. Its spectral content will be arrived at by determining its autocorrelation function* from which the spectral density can be found by using the Wiener-Khinchin theorem, i.e., the autocorrelation function and spectral densities are Fourier transform pairs. Therefore, the signal energy can be related to the articulation score yielding on the average the required scanning waveform to optimize information transmission.

No further work in this area was conducted due to limited funds and time. It appeared more advantageous to continue work using periodic waveforms to scan the speech spectrum.

---

* M. Schwartz, Information Transmission, Modulation and Noise, McGraw-Hill, N.Y., 1959, p. 431.

4.3.2 System Components.

4.3.2.1 Scanning Oscillator.

The scanning oscillator design has a frequency variation of 3 kc which is the bandwidth of the speech spectrum to be compressed. A center frequency of 20 kc for the scanning oscillator was selected because 20 kc filters were readily obtainable and this frequency lies outside the audio band.

The scanning oscillator is a transistorized astable multivibrator and was a preferred design because of greater reliability achieved through a reduced number of components and because of superior stability. Consider the schematic and tuning curve figures 11 and 12. By observing the base return voltage waveform we can see that the cutoff voltage permits alternate conduction of both transistors. The time sequence of these transitions is dependent upon the base resistor, collector to the base capacitor, and base return voltage. Frequency variation is achieved by allowing the base return to seek changes in the voltage asymptote. Referring to the tuning curve figure, note that the frequency excursion is relatively large compared to the small driving voltage used.

Originally a phase shift oscillator figure 13 was designed for the scanning oscillator. Frequency variations are attained by driving non-linear resistances figure 14, with the voltage waveform of the function generator. Since one tube serves as the load in a cascode arrangement,the overall circuit will change resistance as a function of grid voltage at the lower tube. As grid voltage changes, the resulting current produces a change in plate resistance at the upper tube resulting in a non-linear operation. The capacitors and non-linear resistance determine the frequency at which the loop gain is equal to $1 + j0$. Gain is provided by the 12AX7 amplifier in the closed loop to assure that this value of $A\beta$ is maintained for stable oscillation.

Undesirable performance became apparent when it was found, one, that the tuning curve was non-linear, and two, that output amplitude as a function of frequency varied. In an attempt to solve the first problem, three cascoded stages were used. Some improvement was shown but this was insufficient to satisfy our linear requirements. In an attempt to solve the second problem, we used a frequency selective network to control the amplitude over the excursion of frequencies desired. Again, no satisfactory results were obtained for our purposes.

4.3.2.2 Reverberator

Selection of a reverberation unit involves procurement or manufacture of a suitable instrument capable of repeating signals a number of times corresponding to the compression ratio. This in turn, depends on the specifications of the speech compression system. The delay unit should pass the speech frequency spectrum of 250 to 3250 cps with a flat response to within ±3 db. A continuously variable delay of 10 to 100 milliseconds adjustable in 1 millisecond increments is a further requirement. The reverberator should have a characteristic impedance in the order of 1000 ohms, and a transmission loss of 4 to 6 db would be acceptable. However, in order to proceed expeditiously, some of these requirements were waived to procure initial data. Our procedure for this and other items was that if commercial units were available or could be easily modified, then the unit was purchased. If cost, modifications, or delivery did not make this a feasible procedure, then the unit was designed and built.

4.3.2.2.1 Electrical Delay Line

Ultrasonic delay lines of the magnetostrictive (e.g., Ferranti Electric Co.) and piezoelectric (e.g., Bliley Electric Co.) types were rejected, because they required a 100 kc carrier which would complicate the design of the speech compression system.

Electrical networks were then considered and after a similar survey, the ESC Corporation, Palisades Park, N. J., supplied a 3 millisecond delay line for experimentation. Although ESC was ready to build to our specifications, the price was high and delivery uncertain. The 3 millisecond line was tested and data taken, (figure 15).

4.3.2.2.2 Acoustical Delay Line

Acoustical delay lines were investigated and the Fisher Spacexpander procured. The Spacexpander consists of an amplifier following a delay line (electromagnetic transducers) and a potentiometer at the output which controls the decay period after a 33 millisecond delay. The Spacexpander also contains a switch which turned to REVERB ONLY permits the reverberated signal to pass,

while in the MIX position both original and reverberated signals pass. The Spacexpander is manufactured by Hammond Organs, Inc., of Chicago, Ill., who refused to modify the unit and since we were unsuccessful in modifying the Spacexpander, we decided to develop our own delay line. Since the Spacexpander was comparatively inexpensive and readily available, we proceeded to use the Spacexpander for our initial work.

4.3.2.2.3 Test Delay Line

Preliminary experiments were made with round and rectangular bars of brass, copper and aluminum using the test fixture, figure 16a. The following table was considered in choosing the metals and dimensions for test specimens.

TABLE 1
SOUND VELOCITY IN VARIOUS MATERIALS
(Handbook of Chemistry and Physics, 1957, p. 2319)

Material	Sound Velocity (meter/sec)	Sound Velocity (ft/sec)	Length Longitudinal Waves (meter/msec)	Length Shear Waves (meter/msec)
Aluminum	5104	16,740	5	2.5
Silver	2610	8553	2.6	1.3
Brass	3500	11,480	3.5	1.75
Tin	2500	8200	2.5	1.25
Copper	3560	11,670	3.4	1.7
Zinc	3700	12,140	3.7	1.85
Iron & Soft Steel	5000	16,410	5	2.5
Ivory	3013	9886	3.0	1.5
Nickel	4973	16,320	5	2.5

Because initial data showed minute delays, further experiments were made with test fixtures and setups, figure 16 and 17. Test data curves, figure 18, indicate delay and frequency response similar to a comb filter.

Discrepancies in results, (see figure 18), can be ascribed to temperature variations, inconstant stylus positions and pressures, etc. Attempts for consistent results involved dismantling and reassembly of the test fixture and comparison of results, figures 18c, d. Note the use of a 100 k carbon resistor to obtain a flattened curve.

4.3.2.2.4 Magnetic Delay Line

Magnetic recorder types were then investigated and the following firms were consulted regarding magnetic tape reverberators:

American Geloso Electronics, Inc.	New York, N. Y.
Ampex Corp.	Redwood City, Calif.
Amplifier Corp. of America	New York, N. Y.
Audio Master Corp.	New York, N. Y.
Bogen-Presto Co.	Paramus, N. J.
Brush Instruments, Inc.	Cleveland, Ohio
Dictaphone Corp.	New York, N. Y.
Edwards Engineering Co.	Port Washington, N. Y.
Fairchild Recording Equipment Co.	New York, N. Y.
Federal Mfg. & Engineering Co.	Garden City, N. Y.
Kay Electric Co.	Pinebrook, N. J.
Monroe Calculating Machine Co.,Inc.	Orange, N. J.
Telectro Industries, Inc.	Long Island City, N. Y.
Whorf Engineering, Ltd.	Warwickshire, England

The Autovox manufactured by Kay Electric Co., Pinebrook, N. J., was finally selected. In the Autovox, adjusting the position of heads at the periphery of magnetic disks, produces suitable delays. With auxiliary equipment including a Missilyzer recorder for speech spectrographs, we were now able to proceed with our investigation.

4.3.2.3 Narrow Band Pass Filter

A narrow band pass filter was required in the scanning filter system to accomplish bandwidth compression by spectrum sampling. Since no variable filters of this type existed, separate fixed filters were considered. Specifications for four filters are listed in the following table:

TABLE 2

SPECIFICATIONS, NARROW BAND PASS FILTERS

Characteristic	Filter #1	Filter #2	Filter #3	Filter #4
Center frequency	20,000 cps	20,000 cps	20,000 cps	20,000 cps
Bandwidth (3 db attenuation)	1500	1000	750	500
Attenuation	60 db min.	60 db min.	60 db min.	60 db min.
Shape Factor $\left(\frac{\text{Bandwidth 60 db}}{\text{Bandwidth 6 db}}\right)$	2 to 3	2 to 3	2 to 3	2 to 3
Input and Output Impedance	600 ohms	600 ohms	600 ohms	600 ohms
Phase Variation (Linearity with Bandwidth)	10%	10%	10%	10%
Approximate Compression Ratio, $\left(\frac{\text{Speech Spectrum Bandwidth}}{\text{Filter Bandwidth}}\right)$	2:1	3:1	4:1	6:1

L-C filters were investigated and Deeco Instruments of Van Nuys, California, were able to supply filters meeting the filter #1, 2, 3 specifications. Filter #4 was not procurable since this was extremely expensive and difficult to manufacture. Deeco filters BP-492-1500, BP-492-1,000 and BP-492-750, were tested, (figure 19) and used. Although the Deeco filters were satisfactory and were used in the system; however, the imperfection in phase linearity resulted in some dispersion. The effects

of this dispersion constituted frequency distortion when restoring the speech spectrum.

Since no acceptable commercial filter was available previously, a 1500 cps band-pass filter was designed and tested. Theoretical curves, filter schematics and the frequency response of this designed filter are shown in figures 20, 21 and 22.

Initially, consideration was given to the use of mechanical filters as described by Doelz and Hathaway, Electronics, March 1953, and Hathaway and Babcock, IRE Proc., Jan. 1957. The following manufacturers were consulted regarding mechanical filters meeting our specifications:

Collins Radio Co.	Cedar Rapids, Iowa
Burnell & Co.	Pelham Manor, N. Y.
Raytheon Co.	Newton 58, Mass.

It was found that no mechanical filters which would satisfy or approach our requirements were available or could be ordered under reasonable conditions. Though the problem could probably be solved by using a multi-element crystal filter of the type studied by the Hermes Division of Itek Corporation, Cambridge, Mass., these were in the developmental stage and unavailable. The following L-C filter manufacturers were therefore consulted about supplying lumped constant filters to meet our requirements:

Freed Transformer Co.	Brooklyn, N. Y.
Universal Toroid Coil Winding Co.	Irvington, N. Y.
G. B. Electronic Co.	Valley Stream, N. Y.
V. T. C. Corporation	New York, N. Y.
North Hills Electric Co.	Mineola, N. Y.
Deeco Instruments Co.	Van Nuys, Calif.
Raytheon Co.	North Hollywood, Calif.
Magnetic Systems Co.	Monrovia, Calif.

4.3.2.4 Adjustable Band Pass Filter

Another and different filter aspect involved the elimination of speech information being transmitted through the scanning oscillator, to the receiver portion of the system. High-pass filtering was required and the Krohn-Hite Variable Band-Pass Filter Model 310-AB was obtained, although this was not its original application (First Quarterly Report, p. 7).

4.3.2.5 Function Generator

The function generator equipment used to supply a variety of scanning waveforms included the following: Hewlett-Packard Test Oscillator Model 202A supplied the necessary sine-wave and triangular scans. The sawtooth scan was obtained from a Tektronix oscilloscope.

4.4 ARTICULATION MEASUREMENTS

4.4.1 Introduction

The measurements here obtained were articulation scores for bandwidth-sampled, time-sampled, and time-frequency sampled speech. The measurements obtained and the conditions of the articulation testing of the auditors employed are outlined and discussed in this section. Since the essential element of this project was to test the effectiveness of the speech compression system on the articulation scores of auditors, these will be treated enbloc here. Data and other measurements on the electronics of the system employed will be found in other portions of this report.

4.4.2 Articulation Measurement Conditions

Articulation scores may be obtained by a variety of techniques using phonetically balanced word lists.* In this project, a single articulate speaker was first selected for consistency. Several speaker candidates recorded the phonetically balanced word list PB_1 and the articulation scores of five listeners were examined. Since the chosen speaker JM rated highest, JM then recorded the phonetically balanced word lists PB_2 and PB_3. Obtained from the Harvard Psycho-Acoustic Laboratory, each list contains a thousand single syllable words arranged in random order so that no associations can be made by listeners. Listeners MM, BS and MF were similarly chosen from other candidates, because of their high articulation scores. Each listener could adjust the volume for personal preference. To avoid possible association, random groupings of words within these lists were prepared, and as a further check the tests were repeated for purposes of comparison.

---

*James P. Egan, Articulation Testing Methods, The Laryngoscope, vol. 58, pp. 955-991, September 1948.

The phonetically balanced word lists were recorded on a magnetic tape, one-hundred words constituting a run. Bandwidth limiting was used to establish a reference for subsequent testing. The speech signal was then time-sampled with a square-wave configuration adjustable from 12 to 100 milliseconds, with and without reverberation. Similarly, time-frequency sampled testing with saw-tooth, sine-wave, etc. configurations was performed and scores recorded for all runs.

4.4.3 Articulation Measurements With Band Limited Speech

After establishing the conditions for determining articulation scores, we proceeded to measure the articulation scores as speech bandwidth is varied. The test set-up used is shown in the block diagram of figure 23. Also shown in this figure is a frequency response curve showing two bandwidths located at 3 db and 20 db points. To ensure that the information content of speech lies within the passband of the filter, wideband noise was added to the speech signal at a signal-to-noise ratio of 20 db.

The curve of figure 24 shows the relationship of articulation scores with bandwidth. The articulation scores point up the significance of the 3 db bandwidth. More significant is the fact that these scores will serve as a reference for future tests taken with the speech bandwidth compression system. The degree of compression will not be determined by the narrowband filter, but will be dependent upon and related to the articulation scores. For example, if the compression system yields an articulation score of 73 and its bandwidth is 1 kc, the compression ratio is determined by obtaining the effective bandwidth from figure 24 (1.25 kc) and dividing by the bandwidth of the narrow-bandpass filter.

Therefore, the true compression ratio is

$$\frac{1.25\ \text{kc}}{1\ \text{kc}} = 1.25$$

Had the compression ratio been determined by taking bandwidth ratios, the ratio would have been

$$\frac{3\ \text{kc}}{1\ \text{kc}} = 3$$

However, it is obvious that the latter ratio is not significant since there is no comparison of essential equivalents.

4.4.4 Articulation Measurements With Time-Sampled Speech

Articulation scores for time sampled speech together with the test setup are indicated in figures 25 and 26. The time sampling circuit or chopper schematic is shown in figure 27.

In both figures articulation scores were obtained for time-sampled speech with and without reverberation. In figure 25, using the Fisher Spacexpander, articulation measurements were made at different settings -- 0, 50, and 95. These settings control the degree of signal reflection so as to result in repetition of the original input with decreasing amplitude. The overall effect or net result is to lengthen the time the signal exists. Considering this effect in the frequency domain, the response of the Spacexpander will be equivalent to the response of a comb filter.

The results shown in figure 25 shows that the average articulation score was lowered by 8 to 10% for both Spacexpander reverberation settings. It also appears as though a 50 msec scanning period contained the highest articulation score which is the optimum period described in the original proposal.

Figure 26 shows that the articulation score has been improved for all scanning periods using the Autovox as a reverberator. Since this data was encouraging, we decided to use the Autovox in our system.

Other reverberator devices used included a 3 msec delay line (ESC) and a 3 msec delay line (ESC) with a feedback amplifier. Figures 28 and 29 illustrate the test setups and data.

Initially, this data had been taken to observe the effects of a reverberation for high interruption rates. The data indicated that using the 3 msec delay line as the reverberator, with a sampling time of 3 msec, and a scanning period of 6 msec, no improvement on articulation resulted. In fact, using this delay line as a reverberator decreased the articulation to about 10%.

The 3 msec delay line was then used in combination with a feedback amplifier. The feedback amplifier schematic shown in figure 30 uses a 90 k potentiometer to vary the loop again, $A\beta$. Note that the amplifier output is terminated in approximately 600 ohms to match the characteristic impedance of the delay line and that the frequency response of the amplifier shown in figure 31 covers minimum and maximum gain settings. Observing the block diagram of the combined delay line and amplifier, the initial signal is repeated with a decreasing amplitude controlled by loop gain. Here the repeated signal will overlap with the following signal sample, but by adjusting $A\beta$, the amplitude will be sufficiently low not to distort the succeeding signal sample, e.g. using an $A\beta$ of 0.5 improved articulation. Sound spectrographs for different $A\beta$ settings figure 32, pictorially show the degree of reverberation overlap.

4.4.5 Articulation Measurements with Time-Frequency Sampled Speech.

Using the project system, the time-frequency sampling data for compression ratios of two and four are shown in figures 33 and 34. Data was first taken on the compression ratio of two, because if results were favorable at this level we could then proceed to take data for the higher compression ratios. Although, results on the compression ratio of four were satisfactory further investigation of still higher compression ratios was not possible because of time and budget limitations.

Data for compression ratio two is shown in figure 33. We may conclude from this data, that the type of scanning waveform does not influence articulation scores appreciably whether with or without reverberation. In some instances reverberation will improve articulation for a specific waveform and scanning rates. Note that in general the articulation scores with and without reverberation are about 90%, which constitutes fairly good intelligibility for systems of this type. The only disturbing factor encountered was the presence of noise when using the sawtooth scanning waveform. As mentioned previously, this is due to the generation of high order transients inherent in the discontinuities of this waveform.

Articulation scores for the compression ratio of four as shown in figure 34 showed about 10% lower values than scores for the compression ratio of two. One pertinent fact, however, is that reverberation on the average improved articulation scores.

5. OVERALL CONCLUSIONS

Results of the speech bandwidth compression tests show that fairly good articulation scores were obtained. Although, limited compression ratios were used, higher compression ratios would have yielded satisfactory articulation scores although not as high as the ones used in this project. We realize that the components used had limitations, mainly the narrow bandpass filter permitted distortion which did not allow correct restoration of the speech spectrum. This was reflected in the lower articulation scores obtained. Had greater effort been directed toward improving and refining the time synchronization of the system, higher articulation scores would have resulted. This is verified by the experimental data taken with and without the Golay delay line. In fact, the articulation scores continually improved with the continual adjustment of time delay for correct synchronization.

Using reverberators for interpolating between samples resulted in articulation scores highly dependent upon the reverberator used. For example, the Autovox proved to be more effective than all others. This may be attributed to the fact that control was readily available by a number of repetitions and that the repetitions remained fairly constant in intensity, and overlap could be minimized.

The statistical method investigated was to provide a means of determining a periodic scanning waveform. Initially our intention was to make measurements on distribution of zero crossings on infinitely clipped speech. Results shown in figure 10a constitute the extent of our progress. Here, the distributions show the relative frequency of occurrence of particular periods of speech signals. No further extension of this work was possible due to limitations of time and remaining funds on this project.

The system appears to have promise for small compression ratios. Its simplicity makes for a ruggedness and a reliability of performance.

6. RECOMMENDATIONS

Although this system provided data for compression ratios of 4:1, the potential available with extended parameters could have furnished compression ratios as high as 8:1. Modifications required to achieve such a compression ratio would demand further investigation of phase dispersion and/or time synchronization. Considering available commercial equipment, e.g., an additional Autovox and a narrow bandpass filter with linear phase characteristics, the present system could have been extended to the compression ratios mentioned previously. However, because of the time and fiscal limitations of the project, this was not feasible.

As a result of the work on the present system, a system concept employing modified filter scanning techniques has been developed. This concept visualizes a sampling of the speech spectrum by locating two filters to intercept changes of formant frequencies and transmit these changes over a narrow band communication link.

A technical proposal of this system is appended to this report as Appendix A.

7. IDENTIFICATION OF PERSONNEL

The following list includes the names,titles and approximate man-hours of work for all key technical personnel directly assigned to the contract and taking part in this project.

	Hours
Dr. M. J. DiToro, Director of Research (Consultant)	--
Dr. L. J. Castriota, Manager of Engineering (Consultant)	--
A. P. Albanese, Section Head and Project Engineer	874-1/2
N. Rothenberg, Section Head	45
L. Zolotnitsky, Engineer	1666
B. Zivatofsky, Engineer	584
N. Wand, Engineer	127
A. Kiriloff, Engineer	92

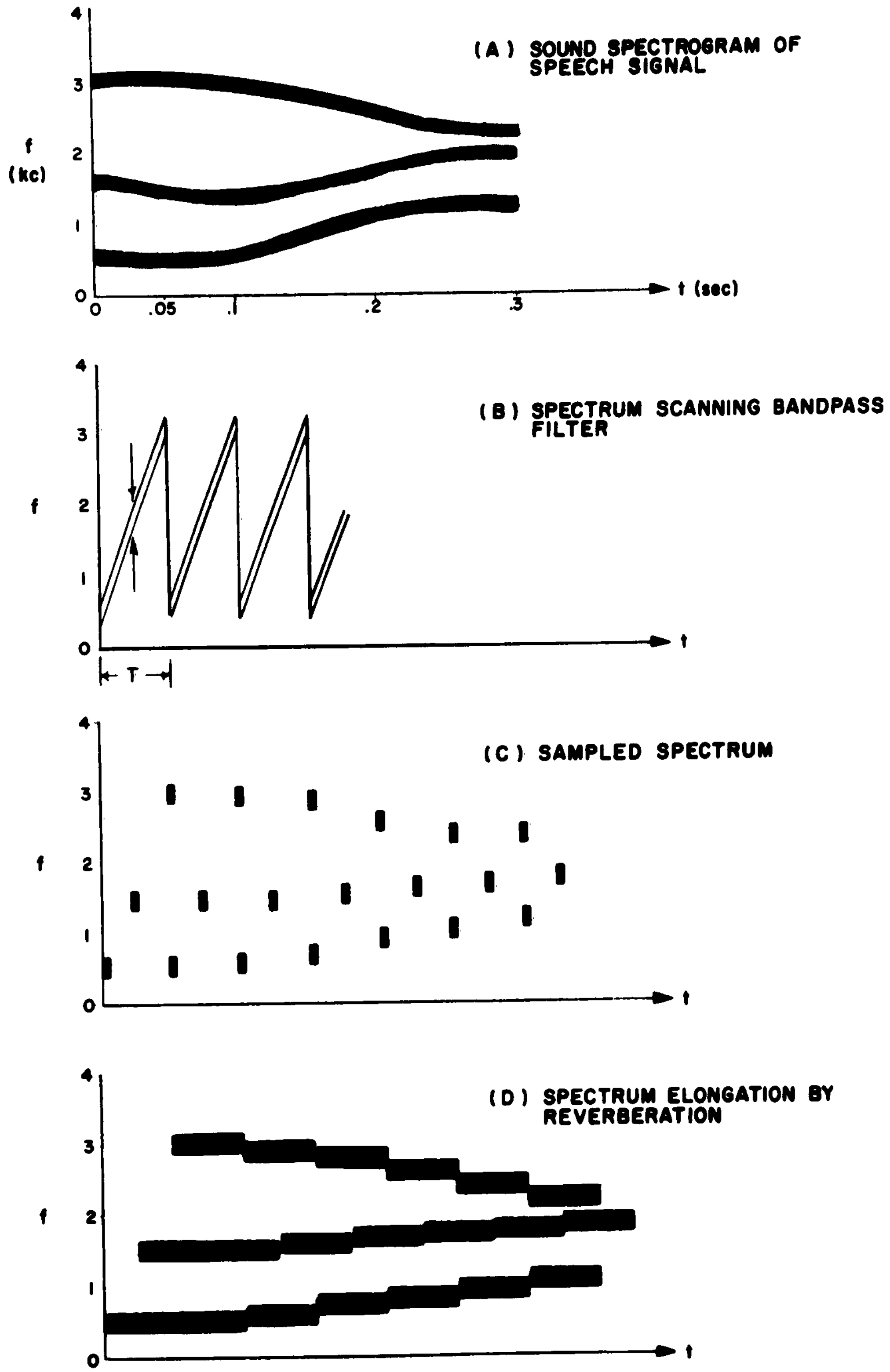
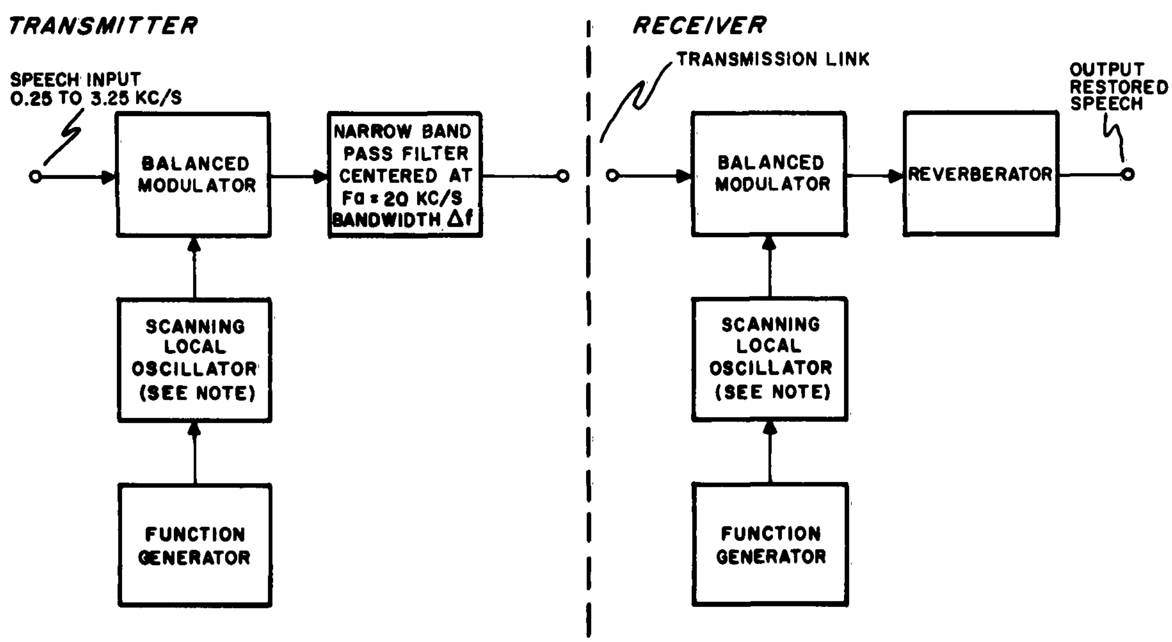


Figure 1. Spectrograms of Time-Frequency Sampling (Sawtooth) with Reverberation



NOTE:

IN AN ACTUAL SYSTEM THE TRANSMITTER SYNCHRONIZATION SIGNAL WOULD BE TRANSLATED TO THE PASSBAND OF THE TRANSMISSION LINK AND DETECTED AT THE RECEIVER BY A FILTER.

Figure 2. Block Diagram of Generalized Speech Compression System

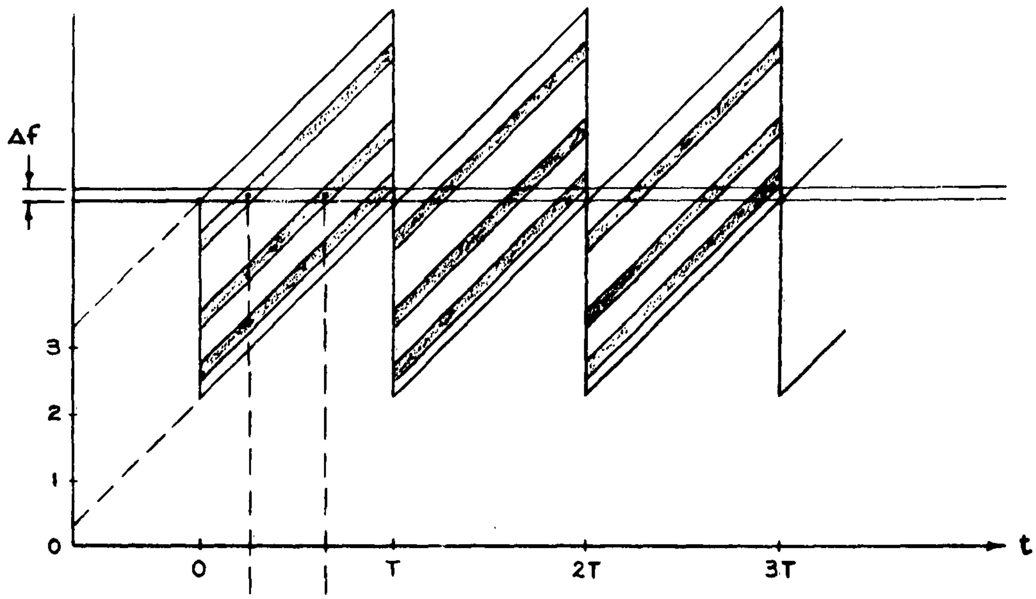


FIG. 3a

FIG. 3b

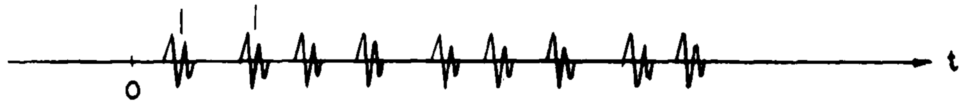


FIG. 3c

Figure 3. Sawtooth Translation of Speech Spectrum

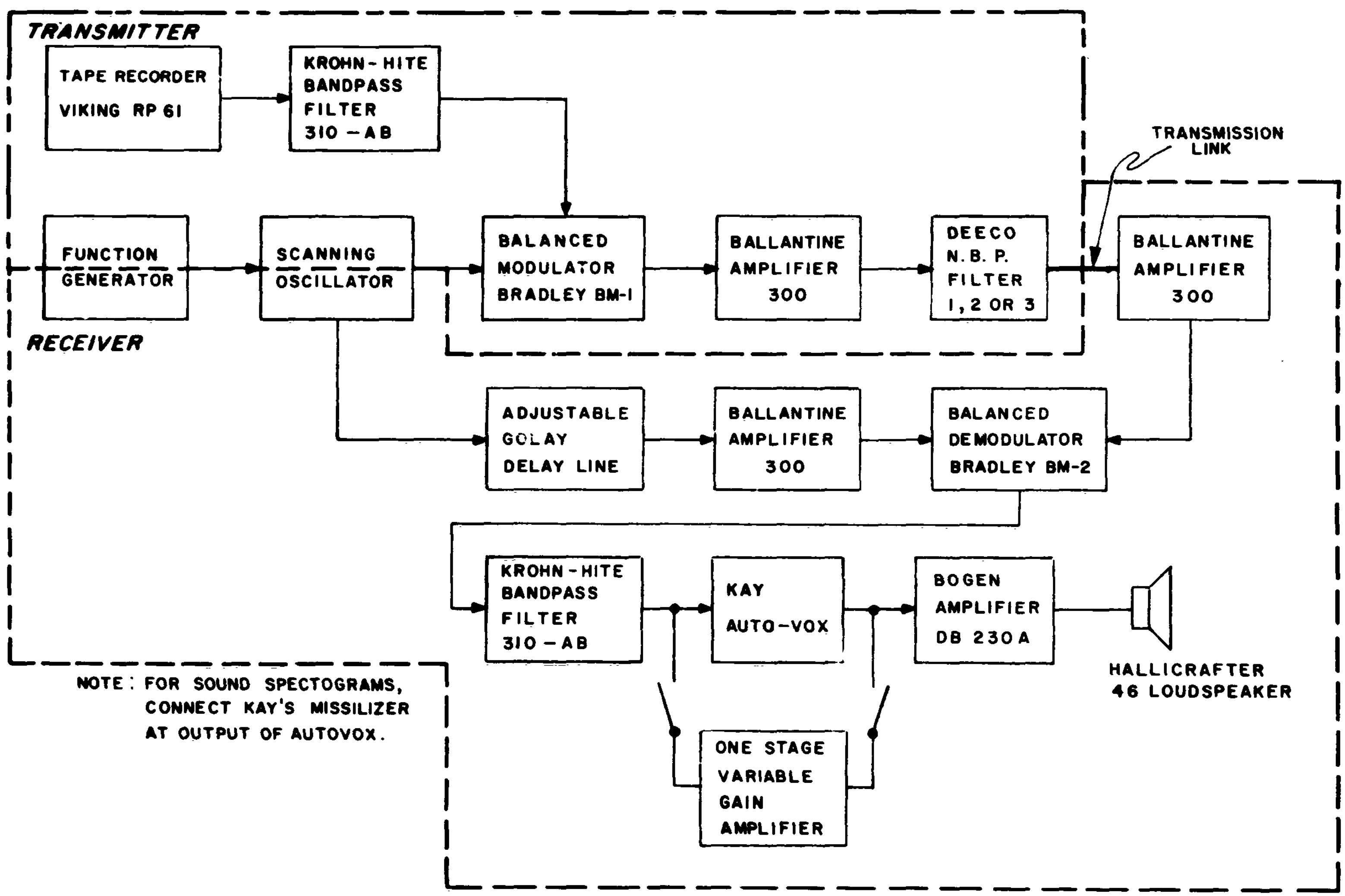
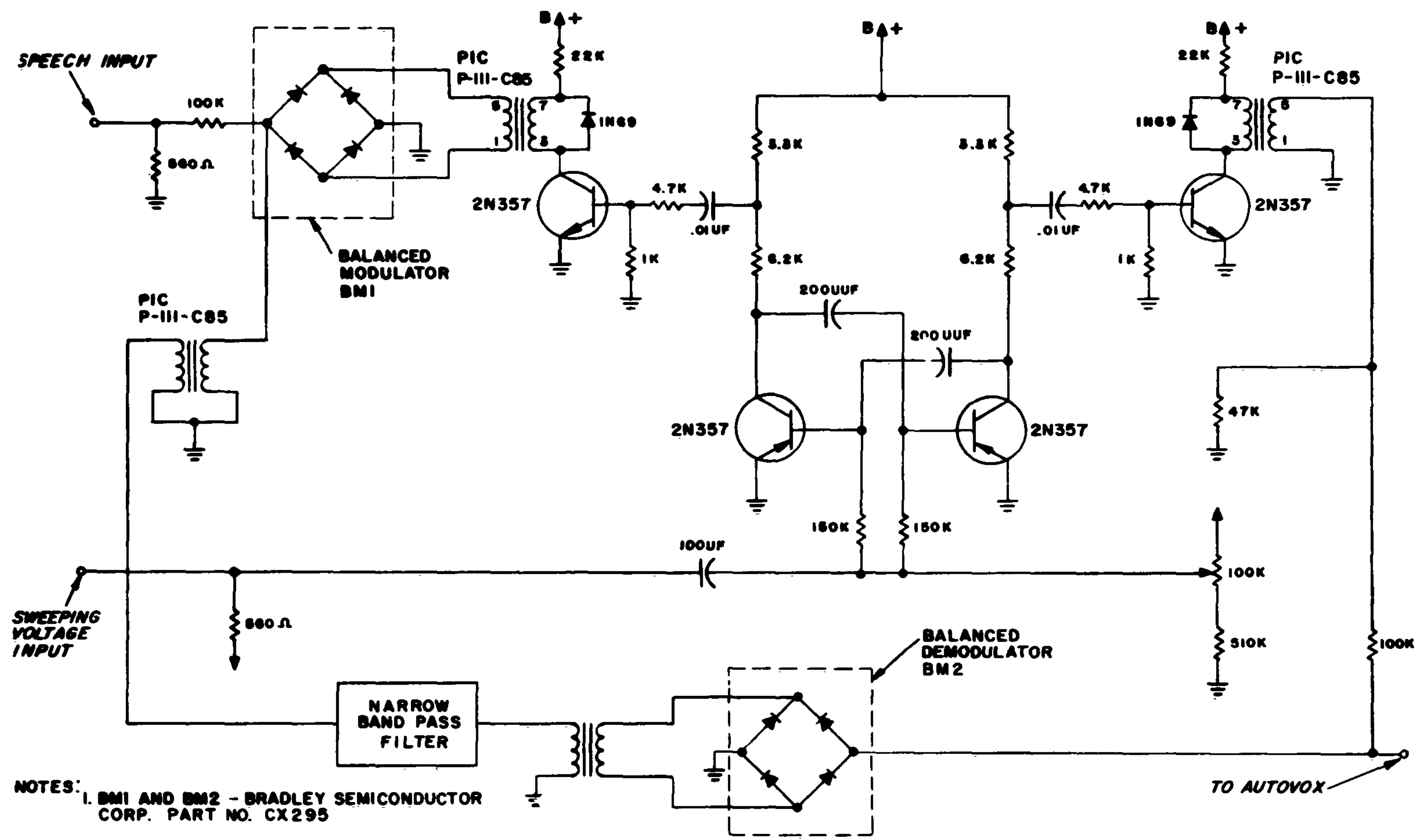
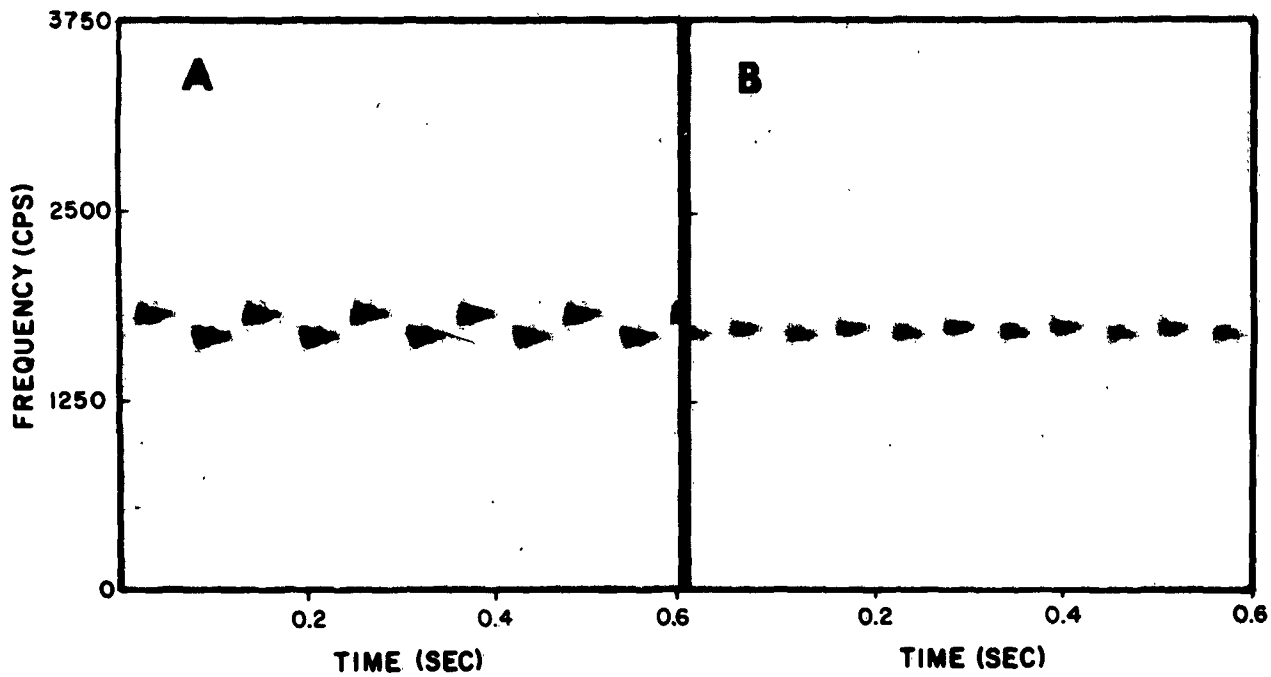


Figure 4. Block Diagram of Project System for Time-Frequency Sampled Speech



NOTES:
1. BM1 AND BM2 - BRADLEY SEMICONDUCTOR CORP. PART NO. CX295
2. T1, T2, T3, T4, - PIC TRANSFORMER CO. PART NO. P111-C85

Figure 5. Schematic for Speech Compression System

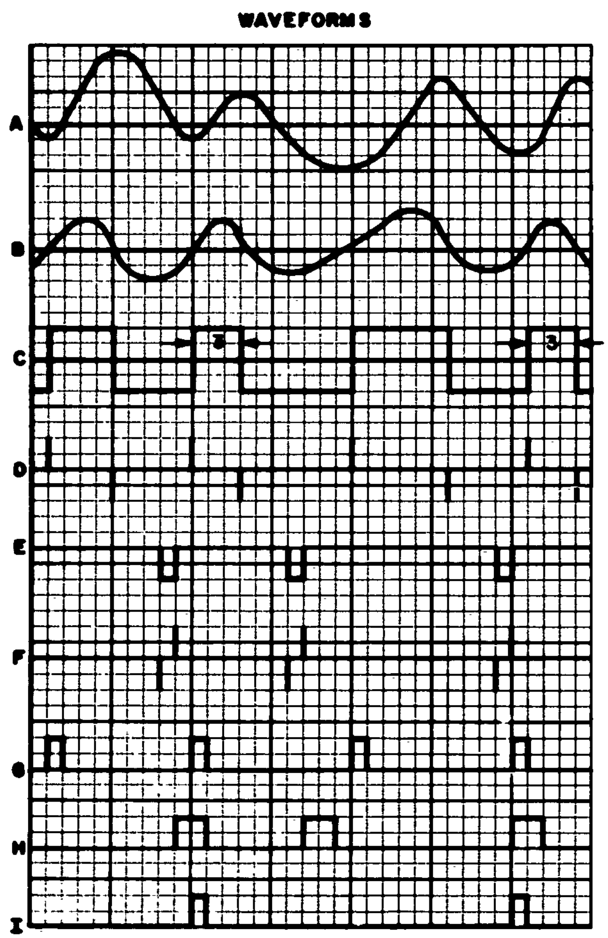


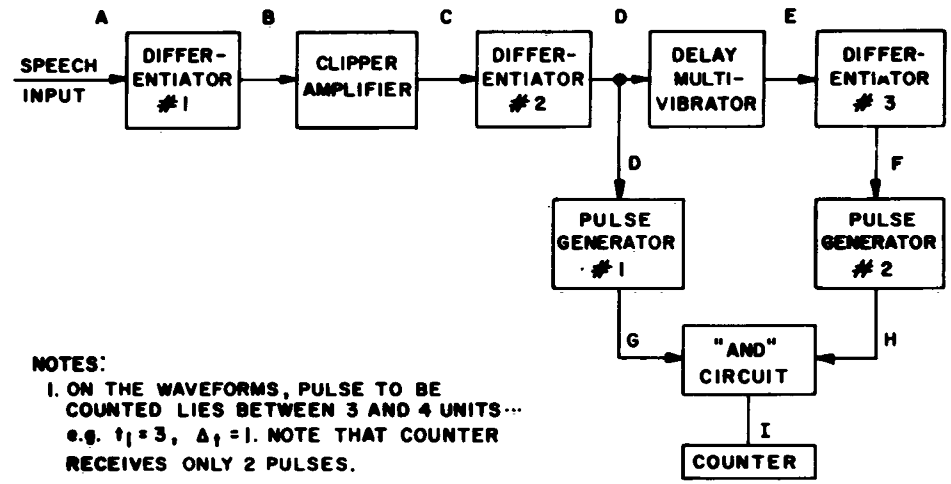
LEGEND:

A. SOUND SPECTROGRAPH, DEECO FILTER 750 CPS BANDWIDTH

B. SOUND SPECTROGRAPH, DEECO FILTER WITH 1.5 MSEC DELAY

Figure 6. Sound Spectrographs for (a) Uncompensated, and (b) Compensated Golay Delay





NOTES:

1. ON THE WAVEFORMS, PULSE TO BE COUNTED LIES BETWEEN 3 AND 4 UNITS… e.g. $t_i = 3$, $\Delta_t = 1$. NOTE THAT COUNTER RECEIVES ONLY 2 PULSES.
2. DISTRIBUTION ANALYZER INCLUDES DIFFERENTIATORS 2, 3, DELAY MULTIVIBRATOR & "AND" CIRCUIT.
3. PULSE GENERATOR 1, 2, ARE HEWLETT PACKARD MODEL NO. 212A, COUNTER IS BERKELEY MODEL NO. 5570

Figure 7. Block Diagram Test Setup for Probability Distribution Analysis

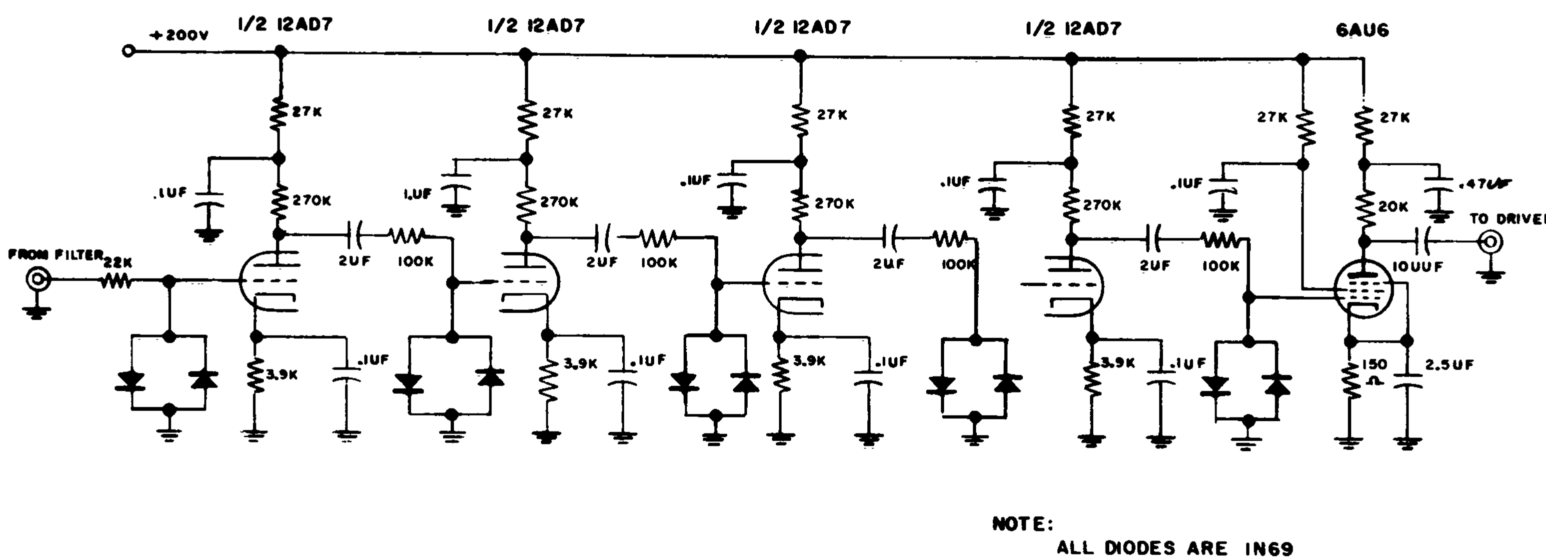


Figure 8. Schematic of Clipper Amplifier

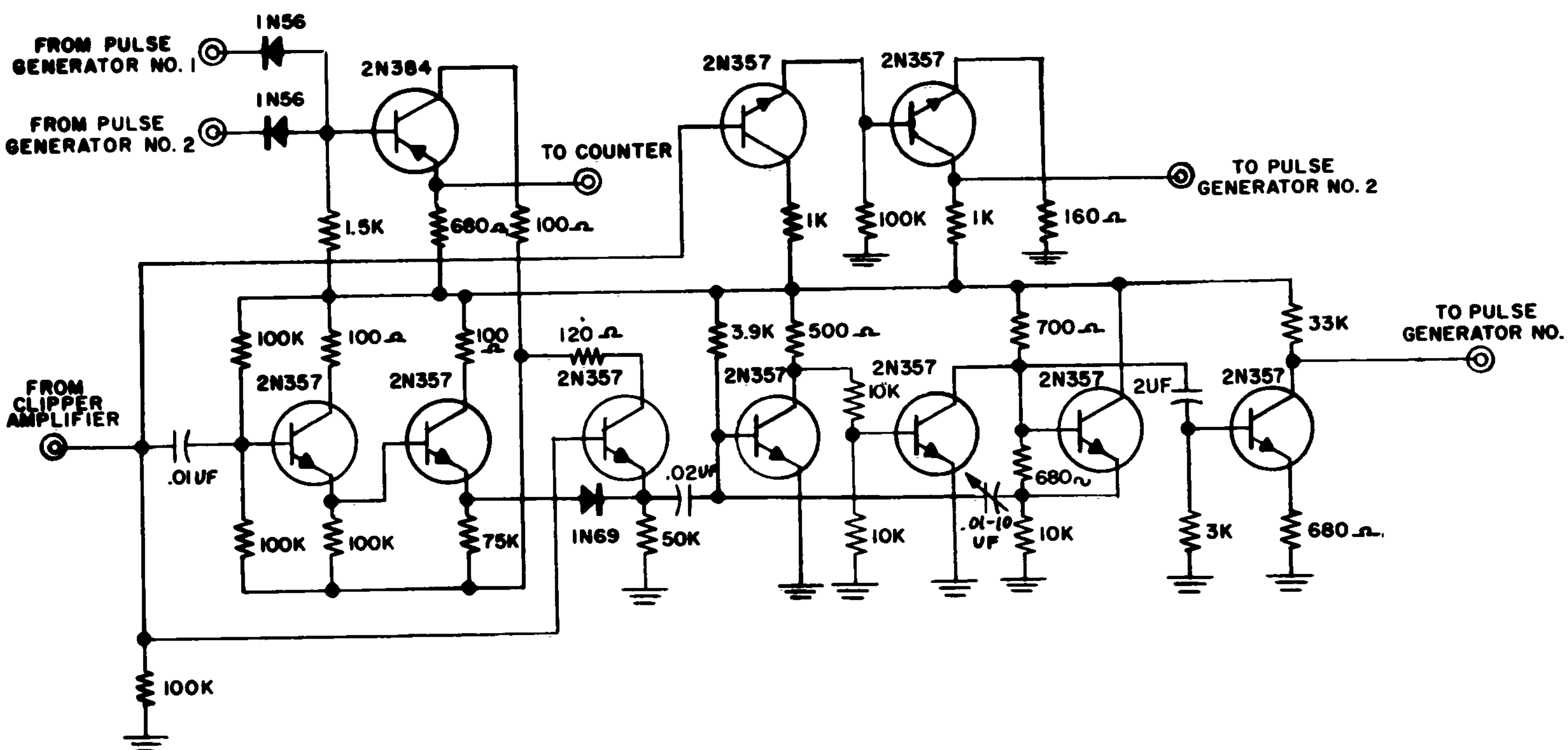
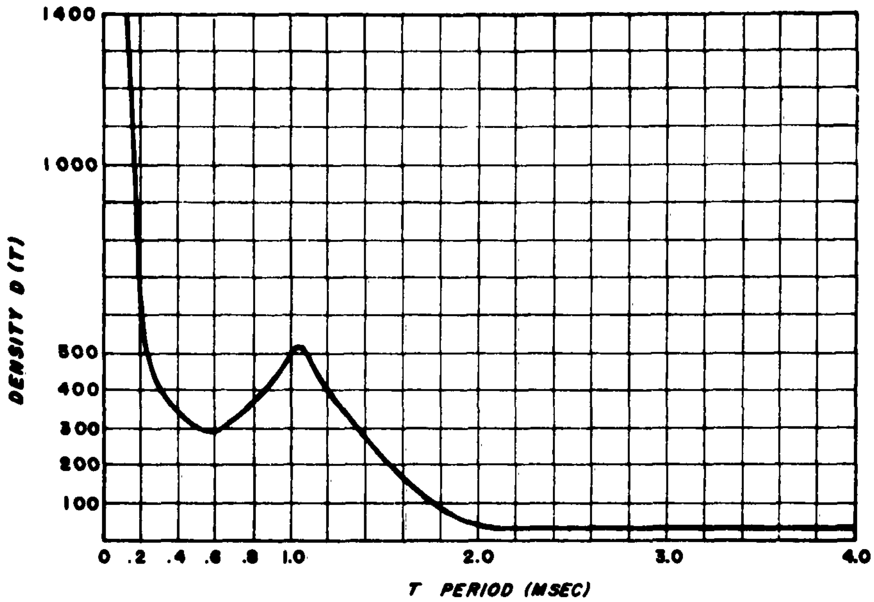
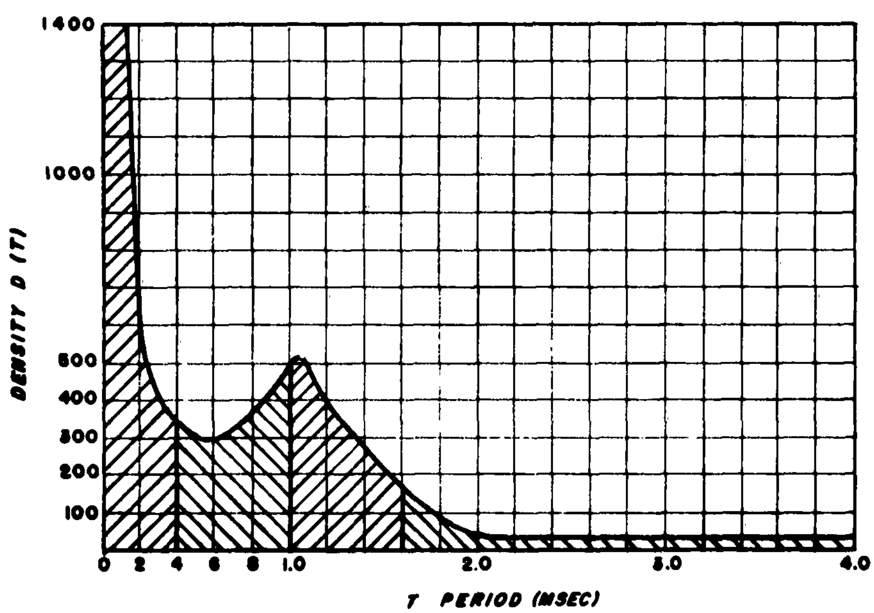


Figure 9. Schematic of Distribution Analyzer



A. Probability Distribution of Zero Crossings for Infinitely Clipped Speech



B. Equal Distribution Areas for Time Periods

Figure 10. Probability Distribution Curves For Speech Zero Crossings

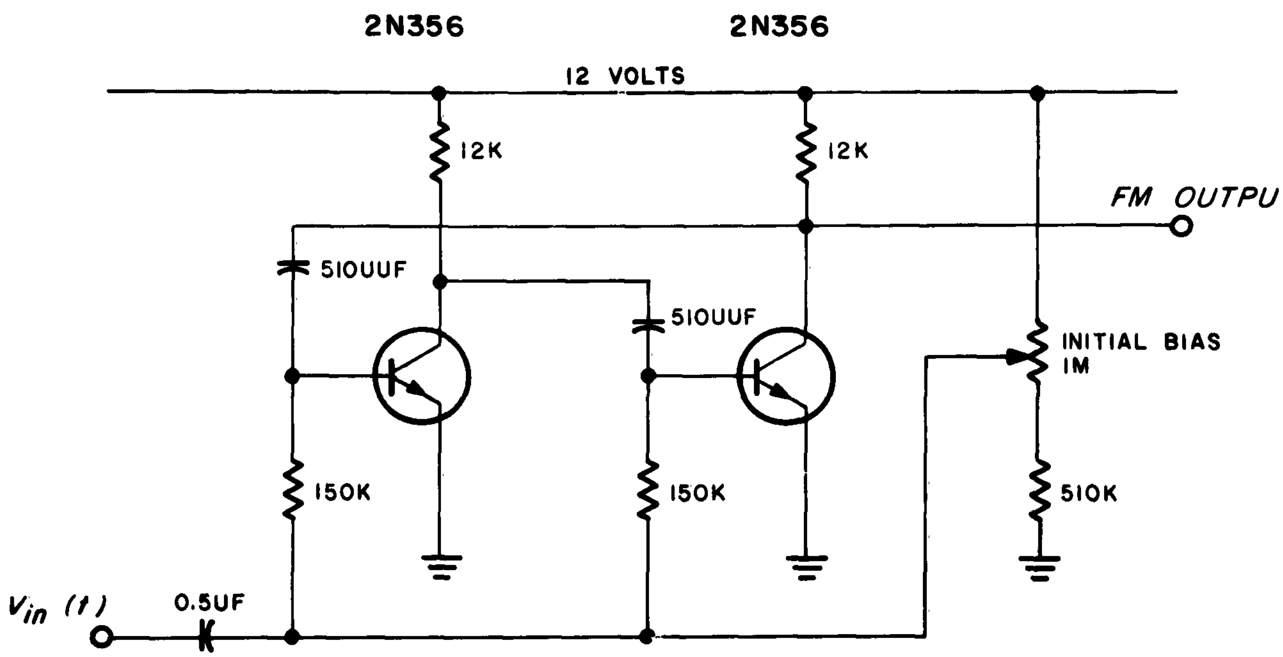


Figure 11. Schematic of Astable Multivibrator Used in the Scanning Oscillator

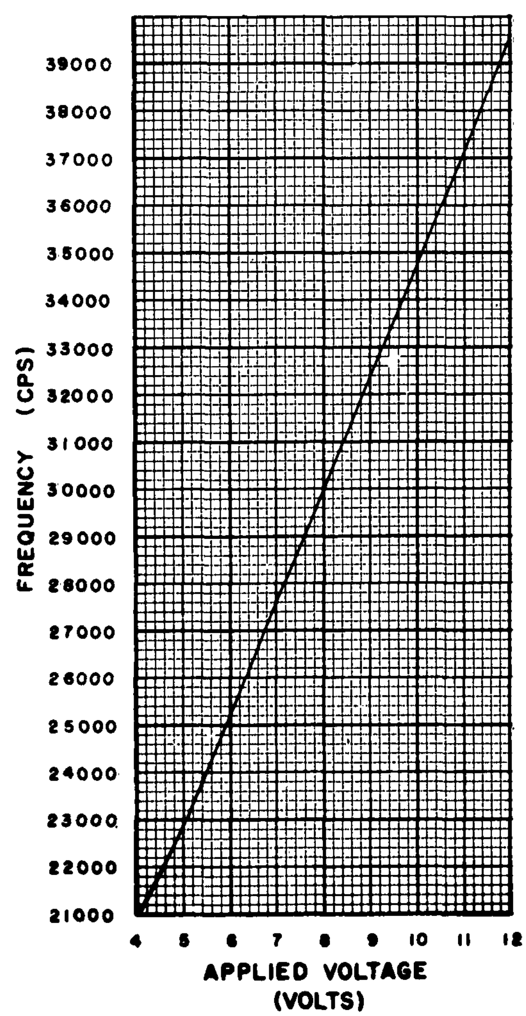


Figure 12. Frequency Variation of Scanning Oscillator vs Applied Voltage

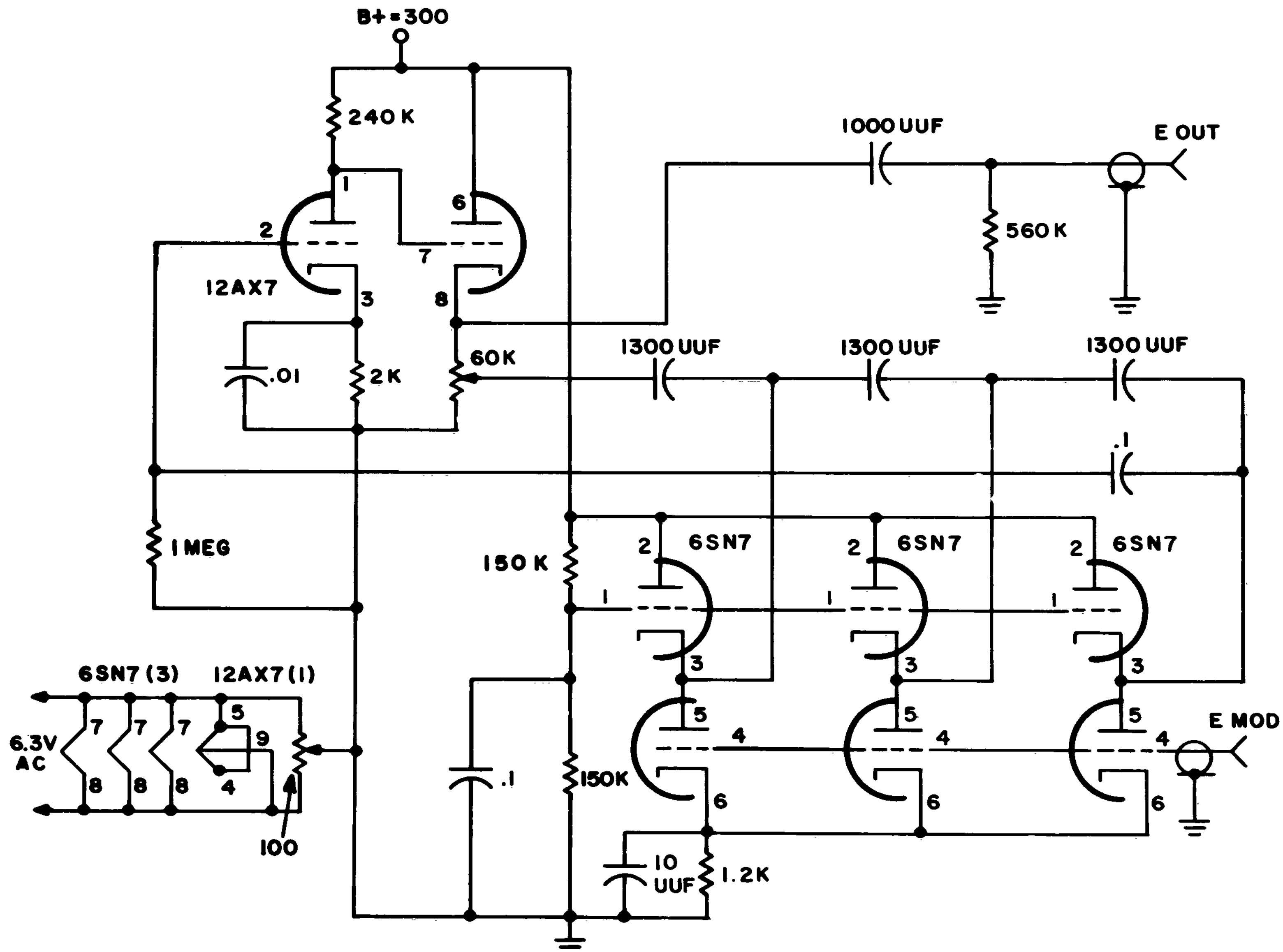


Figure 13. Schematic of Phase Shift Oscillator

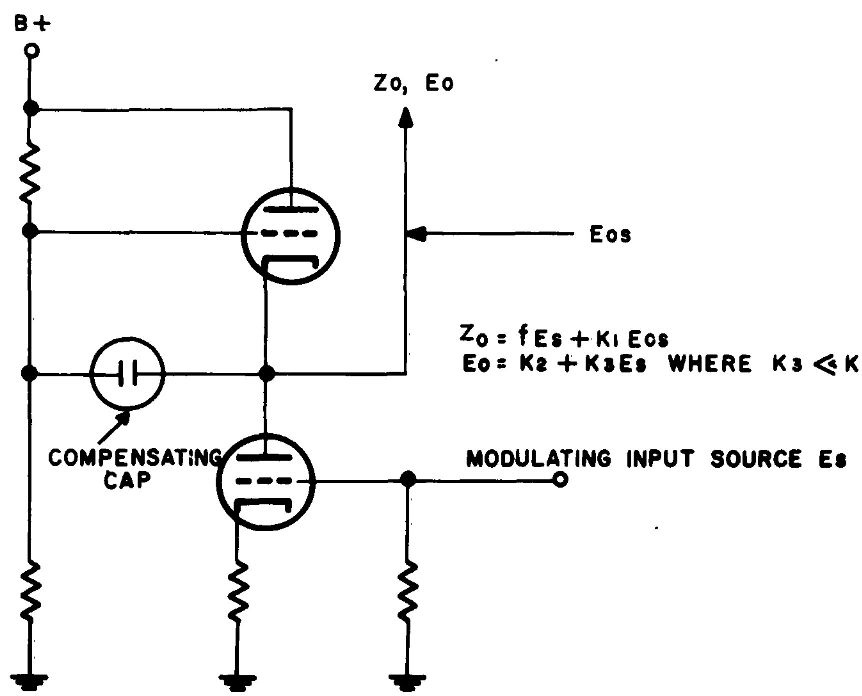


Figure 14. Variable Resistance Network in Phase Shift Oscillator

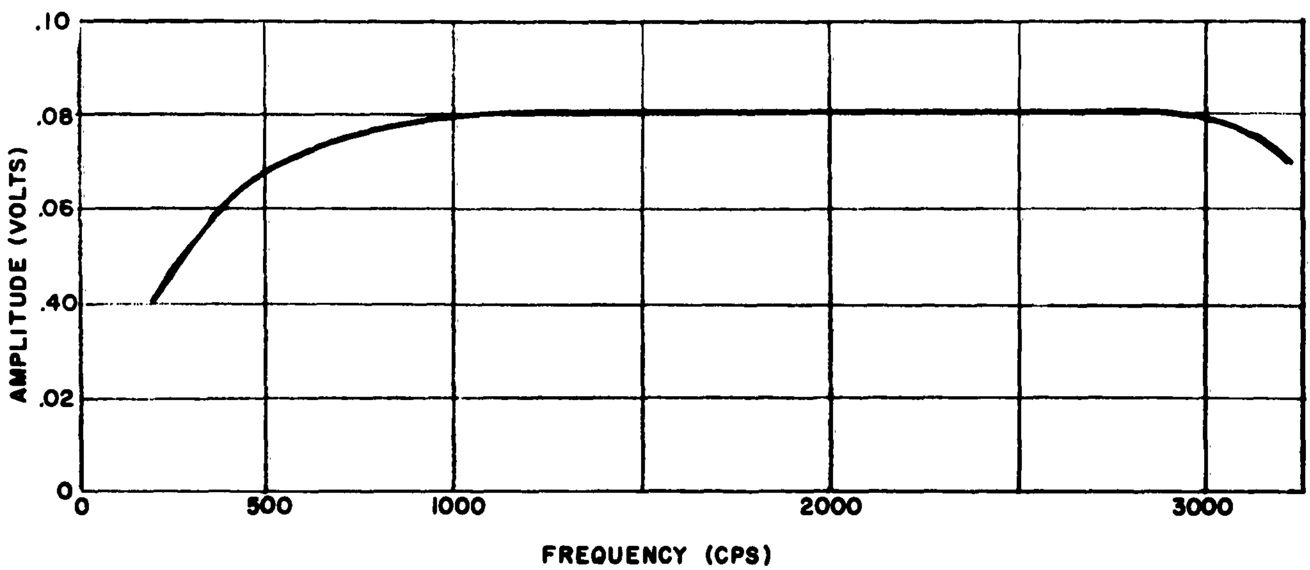
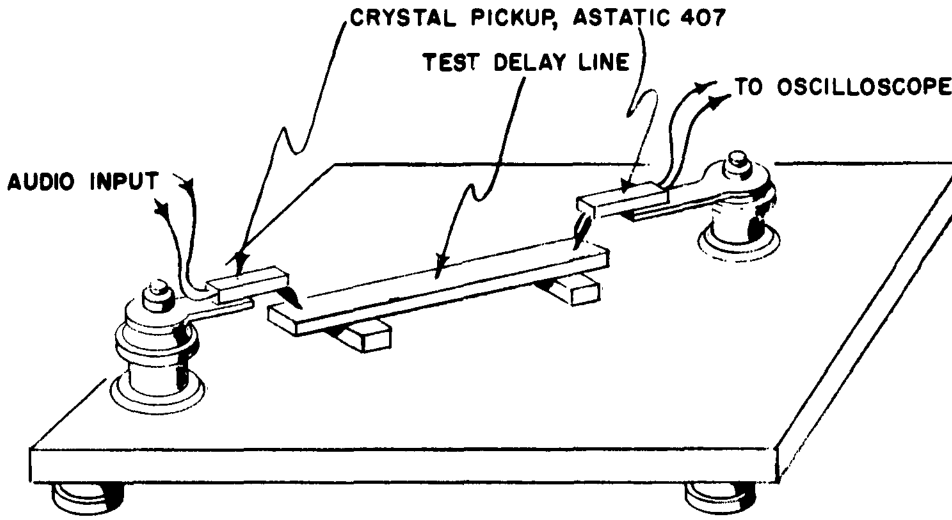
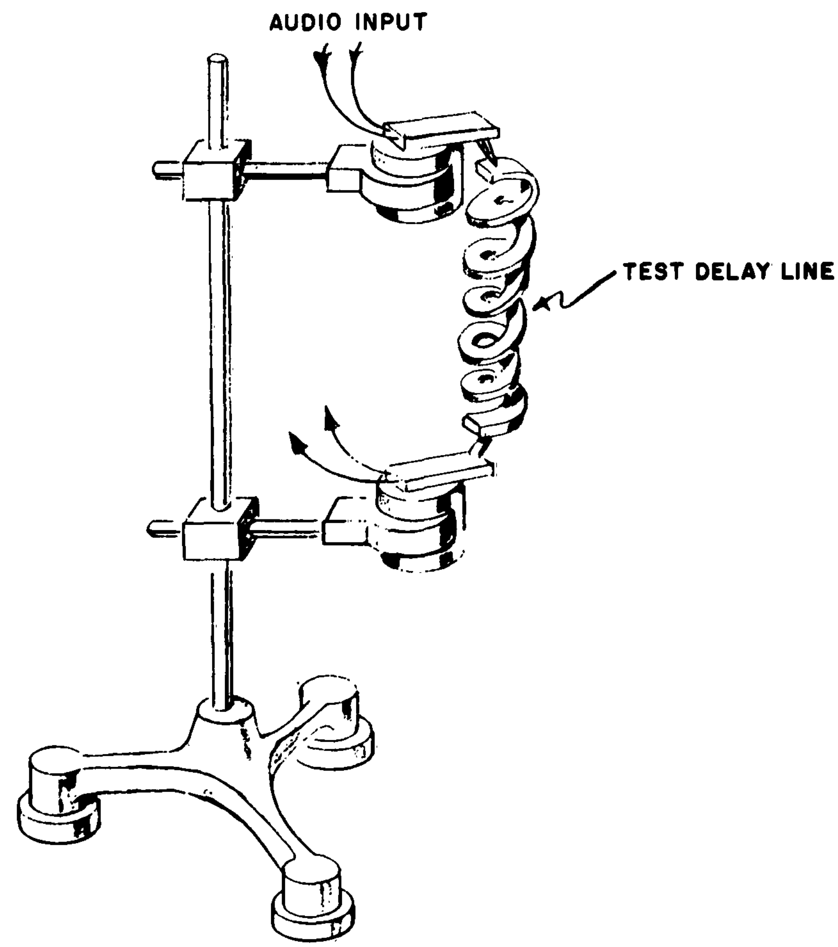


Figure 15. Frequency Response of 3 msec Delay Line (ESC Corp.)

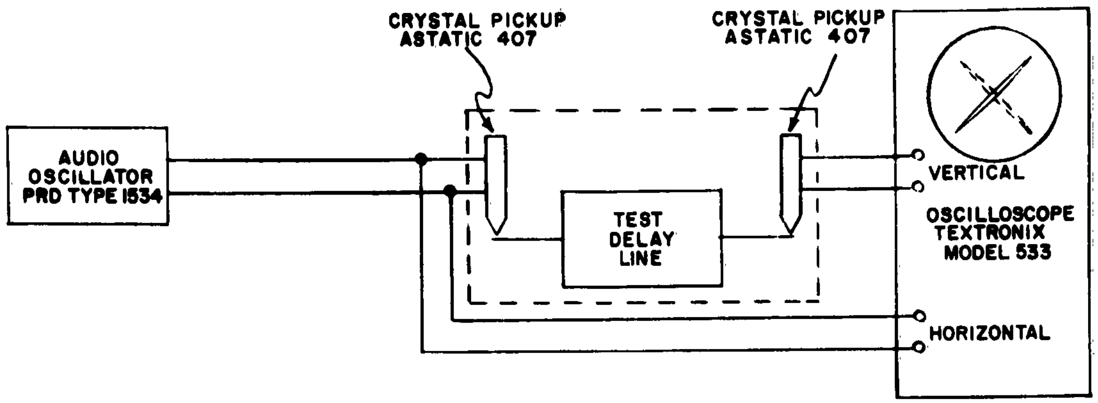


A. Test Fixture For Delay Specimen

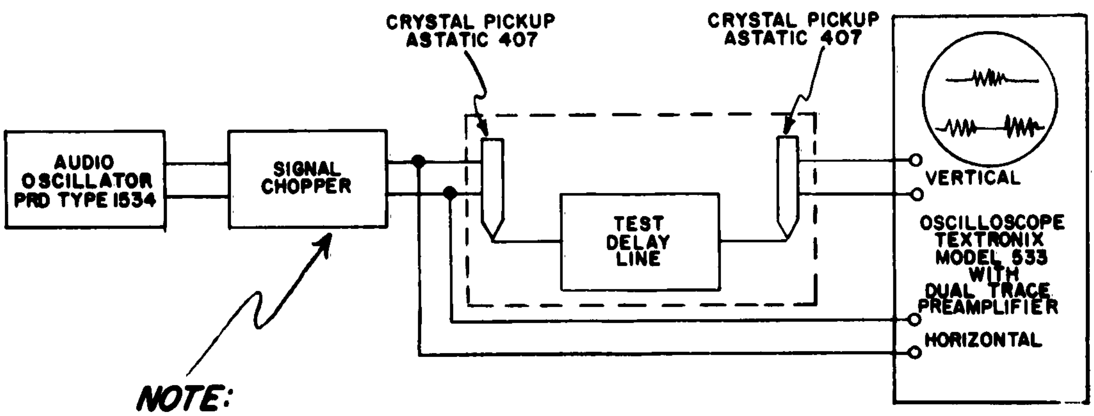


B. Test Fixture of Delay Line

Figure 16. Test Delay Line Apparatus



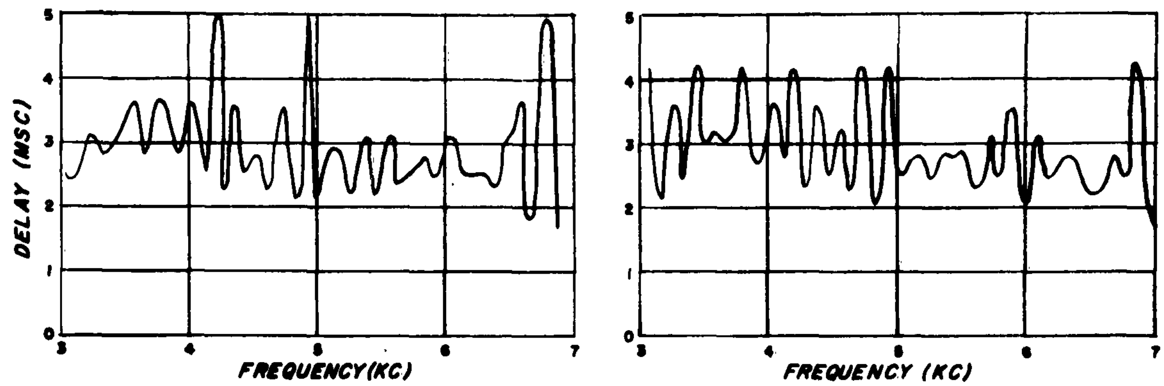
(A) PHASE SHIFT TEST SETUP



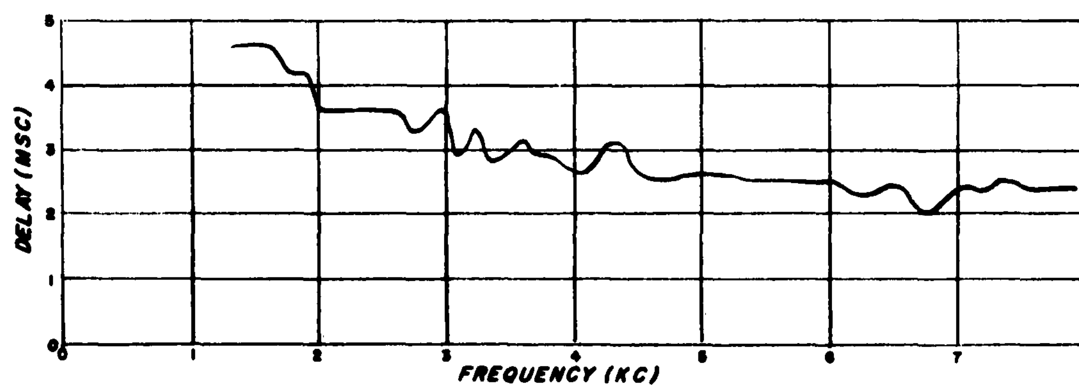
NOTE:
FOR SIGNAL CHOPPER SCHEMATIC SEE FIGURE 27.

(B) SIGNAL SHIFT TEST SETUP

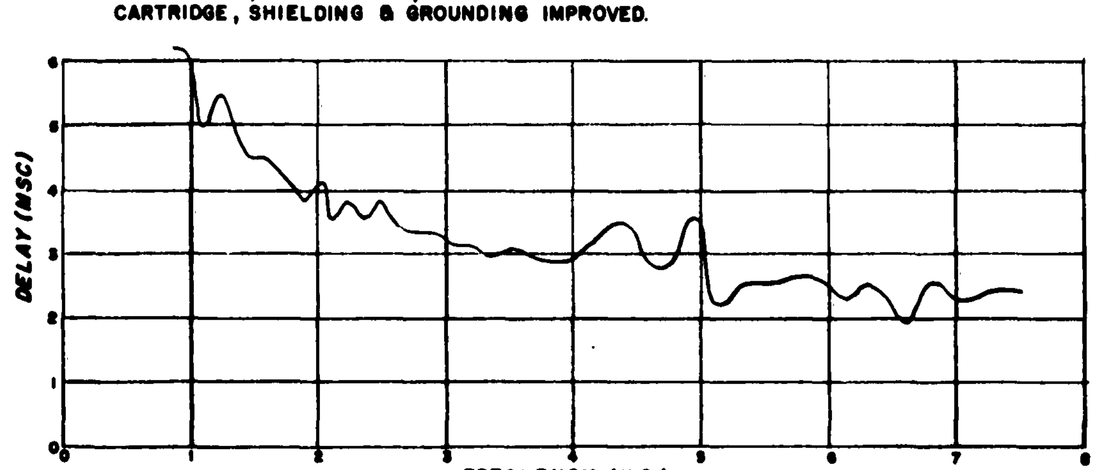
Figure 17. Block Diagram of Test Setups for Investigation of Test Delay Lines



(A),(B) COPPER WIRE HELIX, 18 TURNS, 0.052 IN. DIA, 45 IN. LG, SIGNAL INPUT 0.86V



(C) SAME HELIX, 2.2V SIGNAL, 100K CARBON RESISTOR IN PARALLEL WITH OUTPUT CARTRIDGE, SHIELDING & GROUNDING IMPROVED.



(D) SAME TEST FIXTURE (C) AFTER DISMANTLING & REASSEMBLY.

Figure 18. Test Delay Line Data Curves

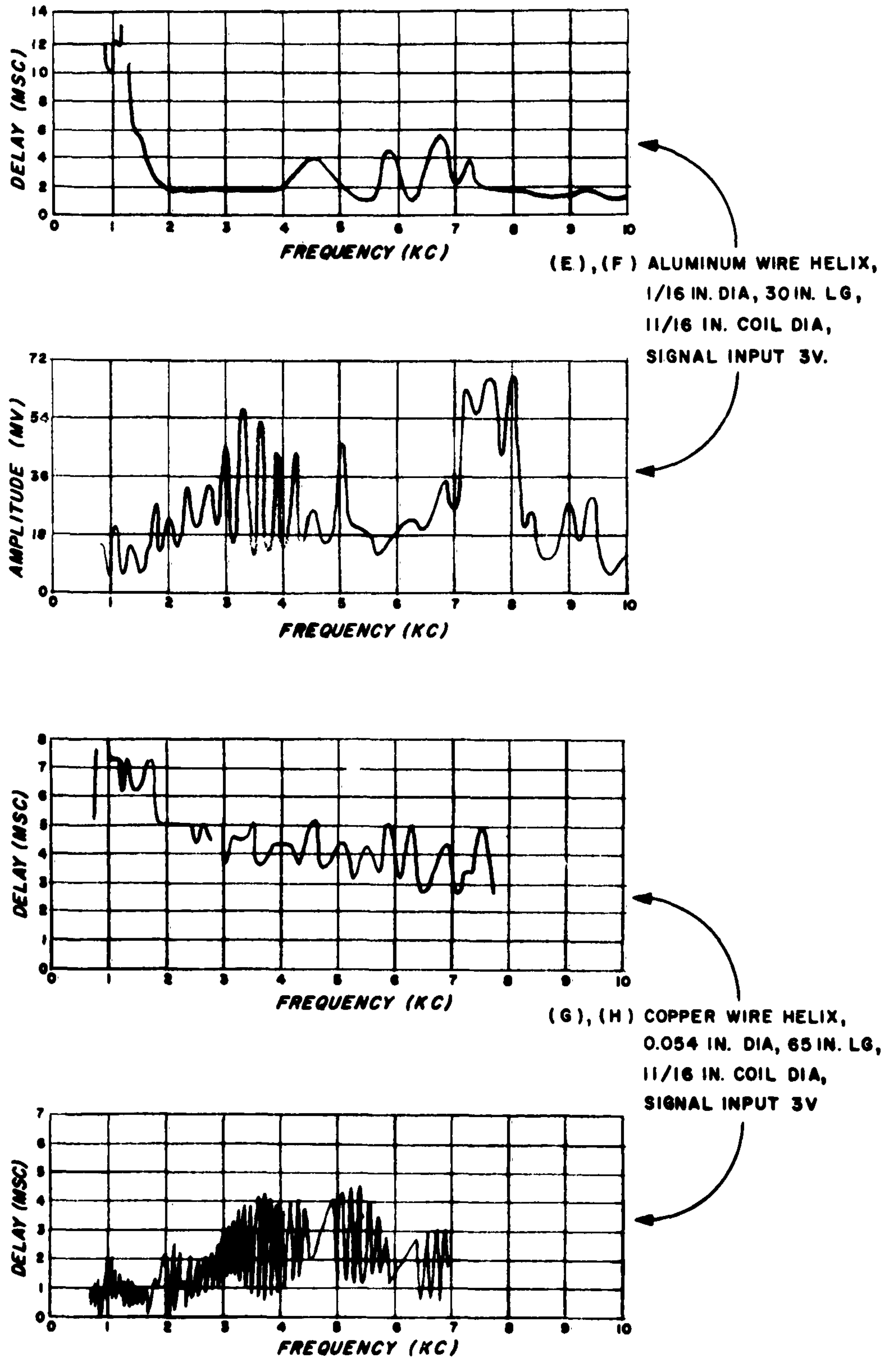


Figure 18. (continued) Test Delay Line Data Curves

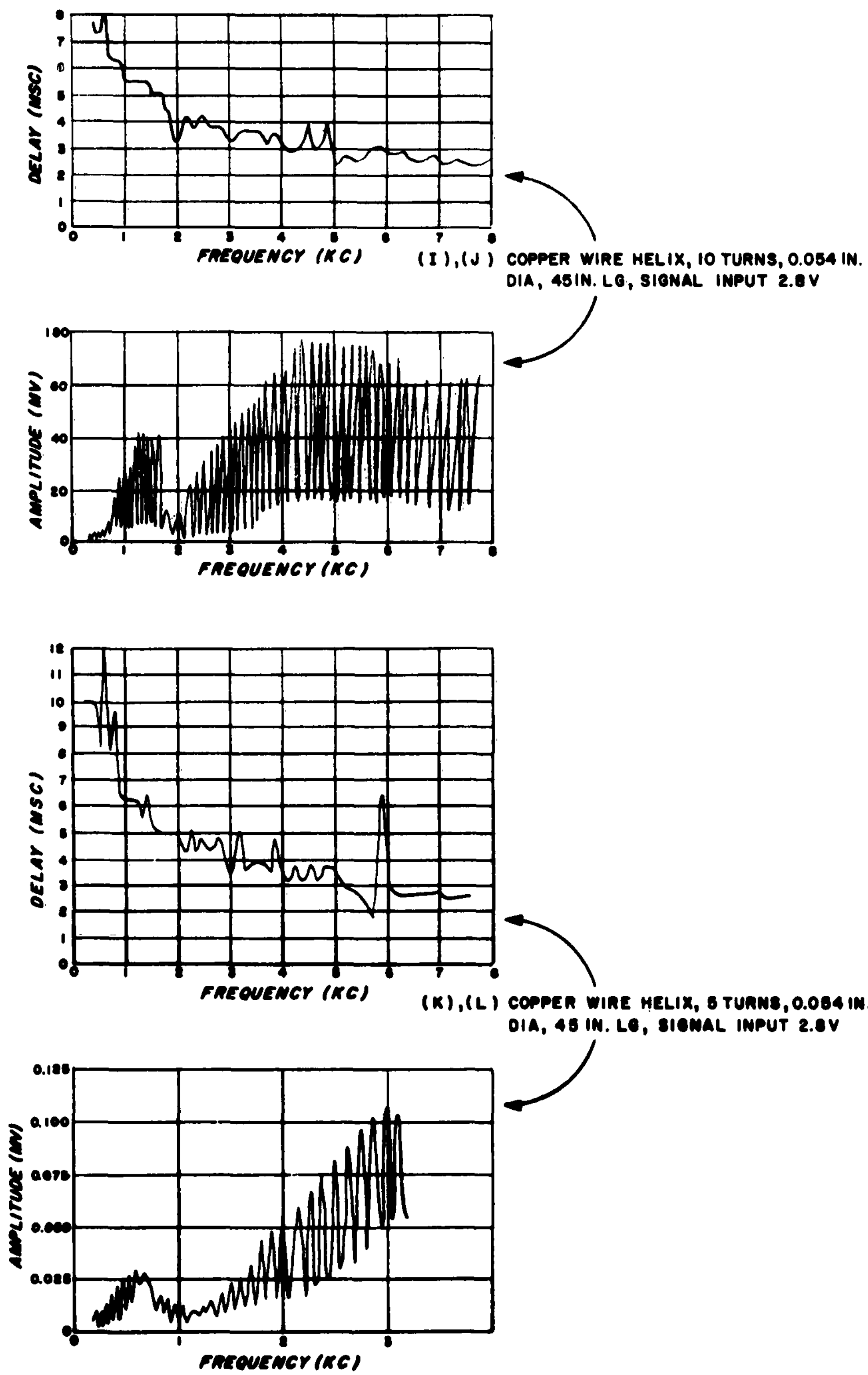


Figure 18. (continued) Test Delay Line Data Curves

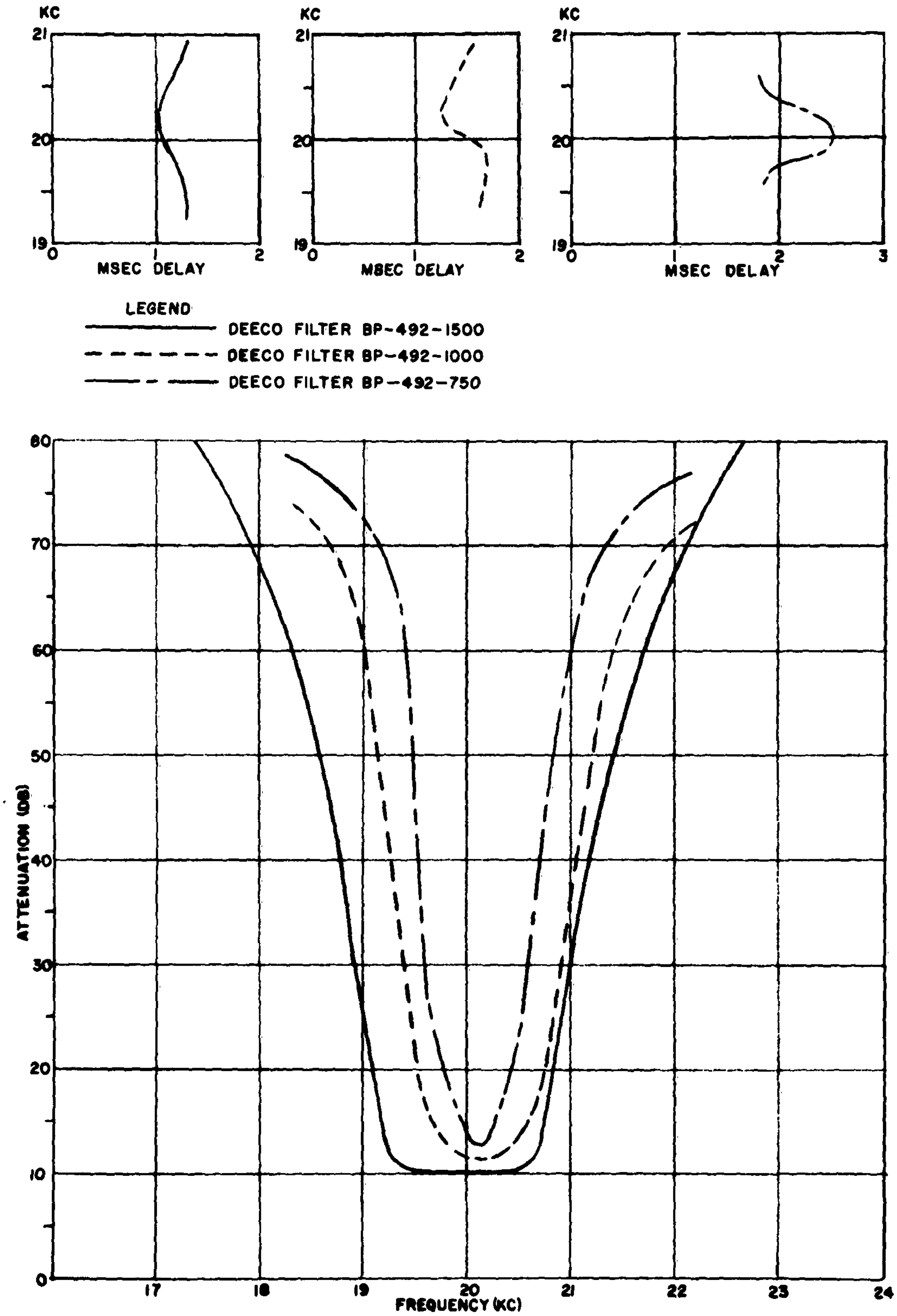


Figure 19. Deeco Filter Data Curves

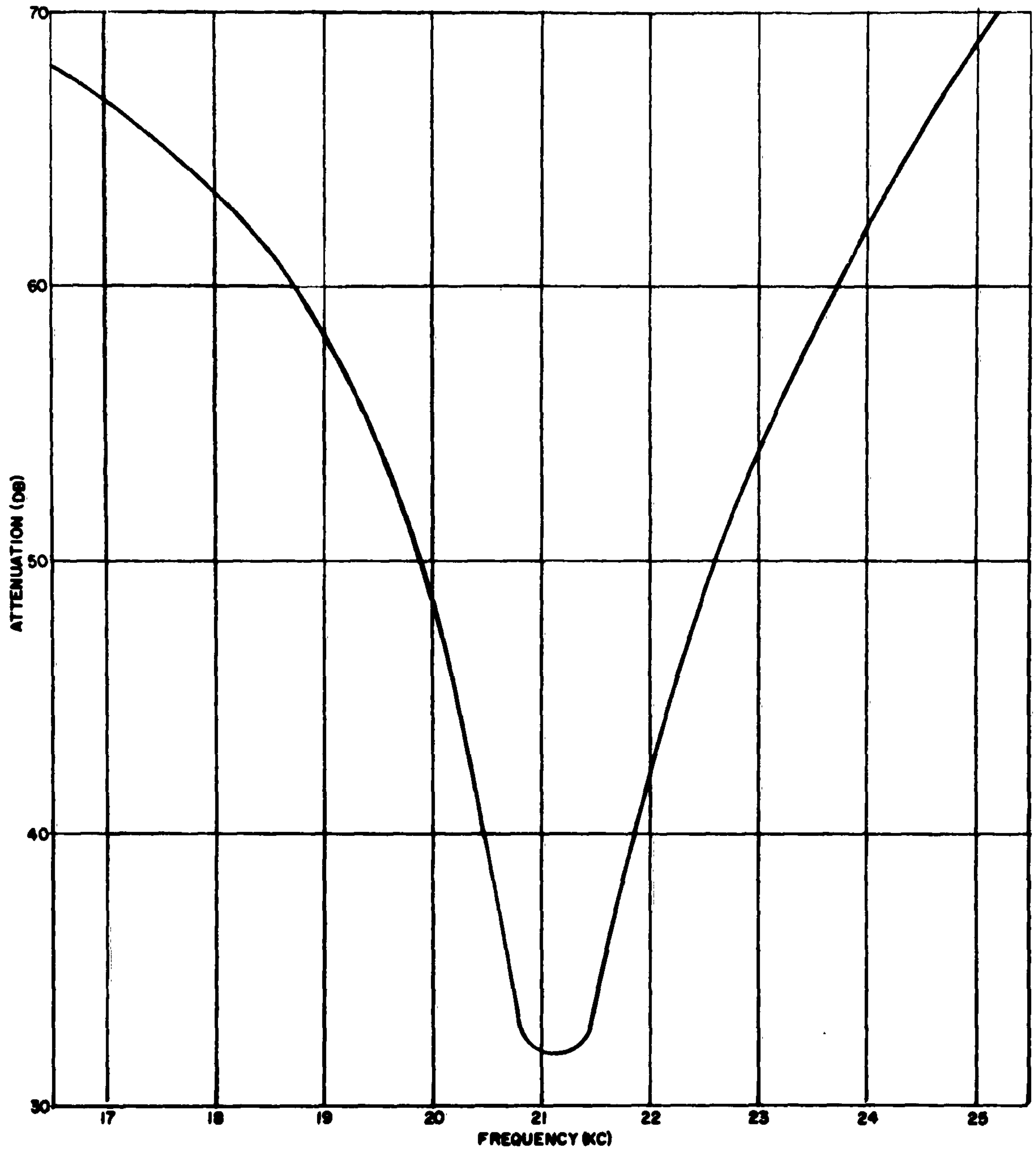
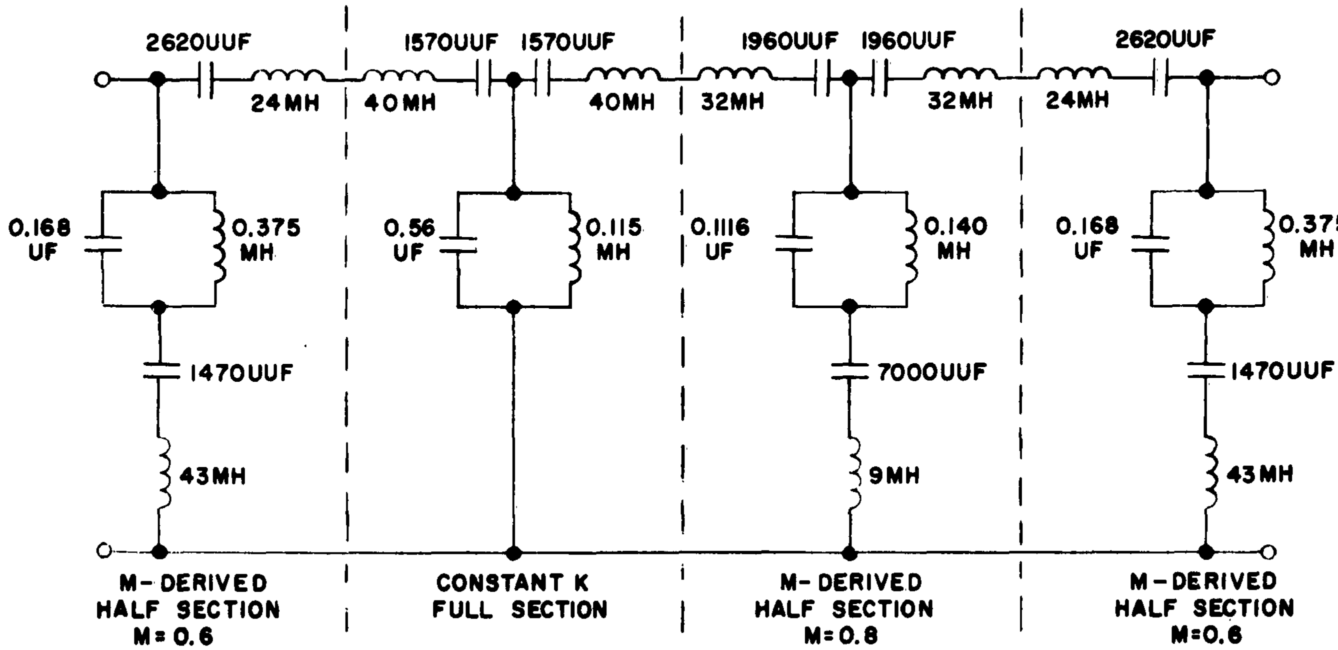
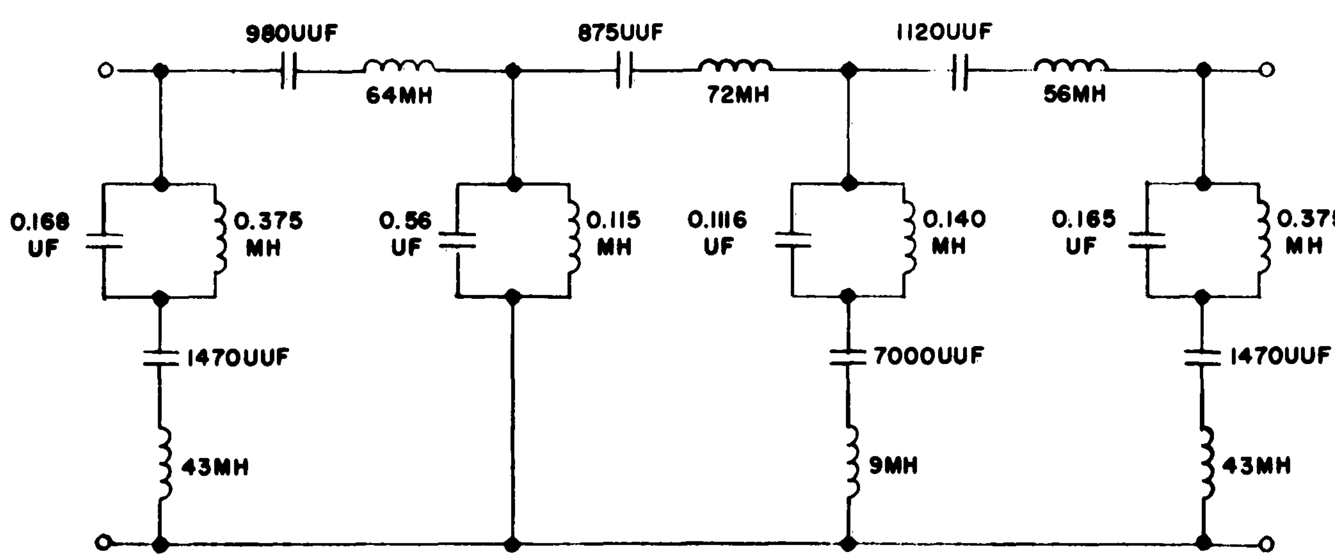


Figure 20. Frequency Response of Designed Narrow Band Pass Filter (1500 cps)



A. CALCULATED NETWORK SHOWING INDIVIDUAL FILTER SECTIONS



B. FILTER SCHEMATIC WITH CALCULATED VALUES

Figure 21. Designed Narrow Band Pass Filter Schematics (1500 cps)

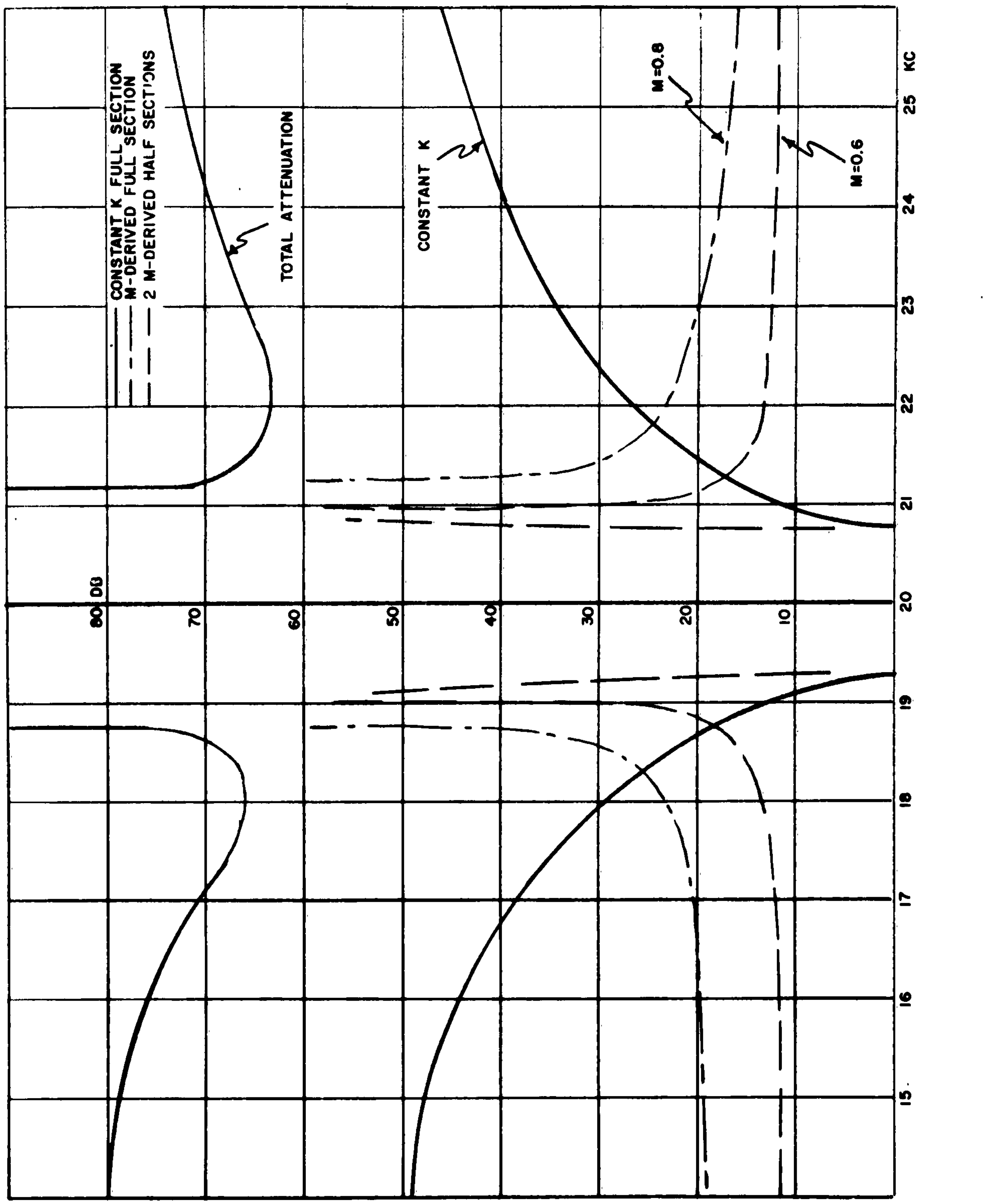


Figure 22. Calculated Response Curves of Designed Narrow Band Pass Filter

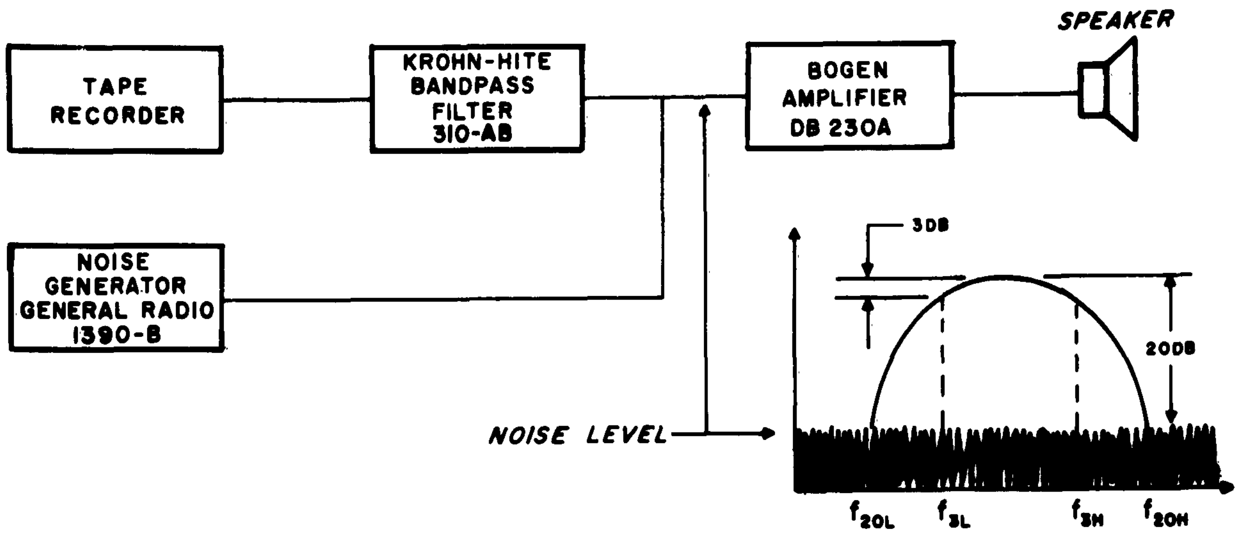


Figure 23. Block Diagram For Articulation Measurement of Band-Limited Speech

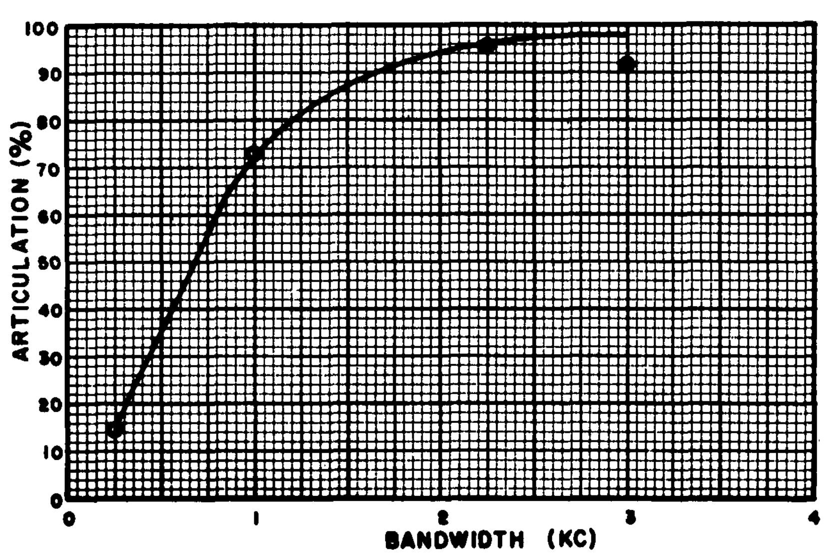


Figure 24. Articulation vs Bandwidth Curve

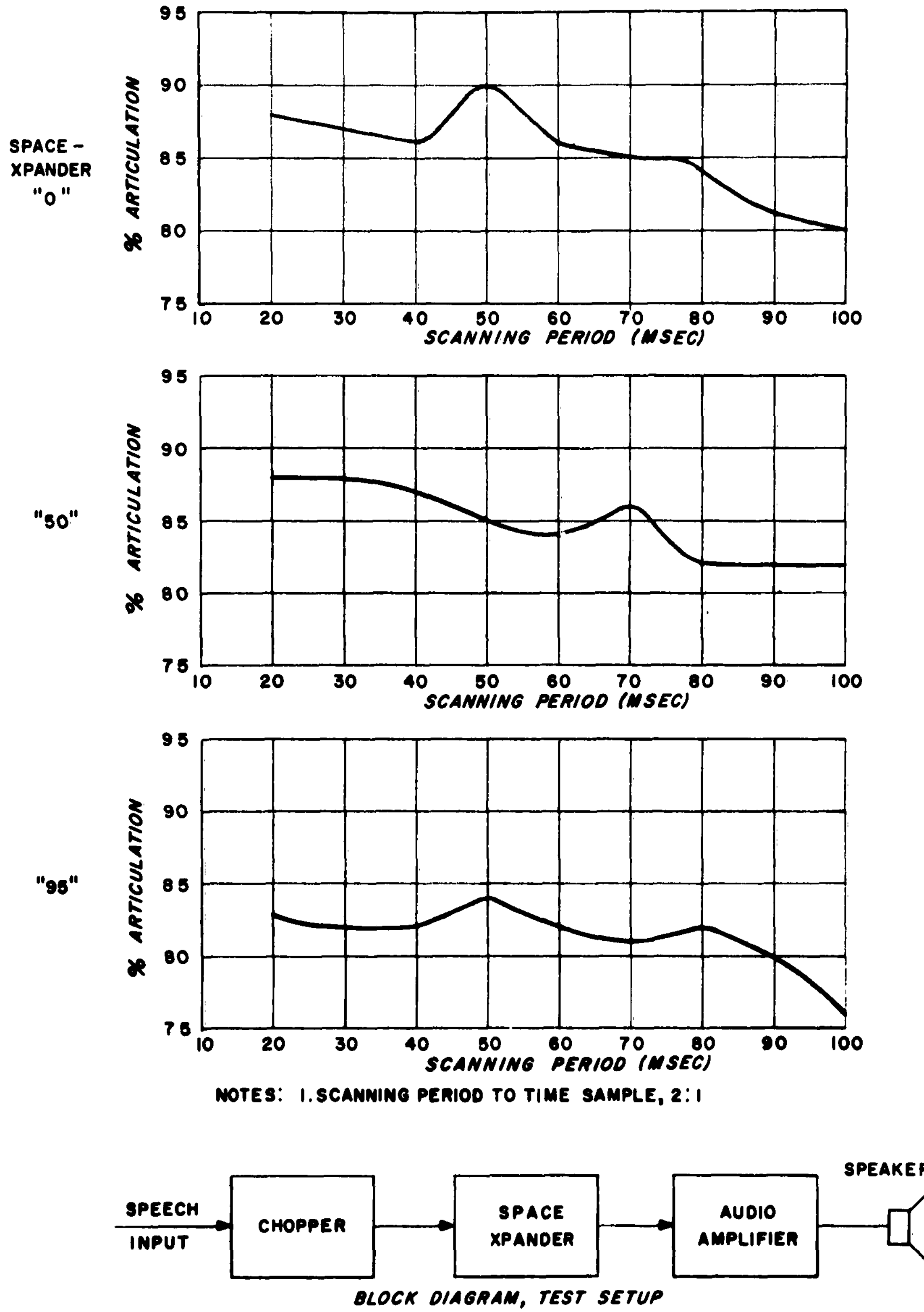
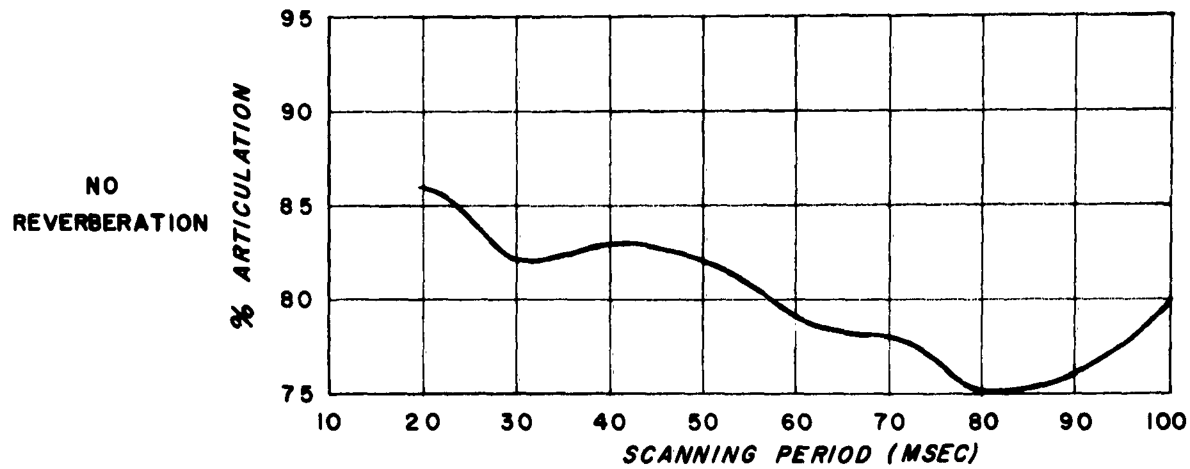
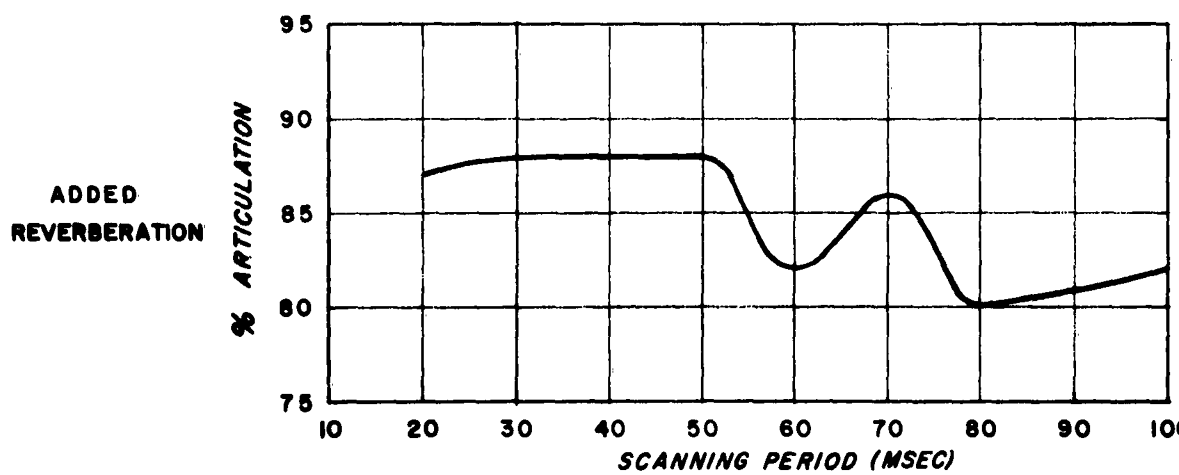
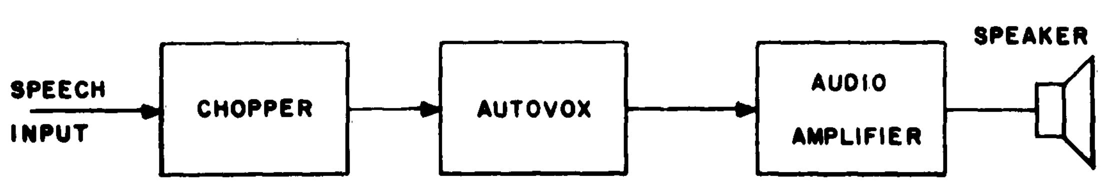


Figure 25. Articulation Measurements for Time-Sampled Speech Using Spacexpander Reverberator





NOTE: SCANNING PERIOD TO TIME SAMPLE, 2:1



BLOCK DIAGRAM, TEST SETUP

Figure 26. Articulation Measurements for Time-Sampled Speech using Autovox Reverberator

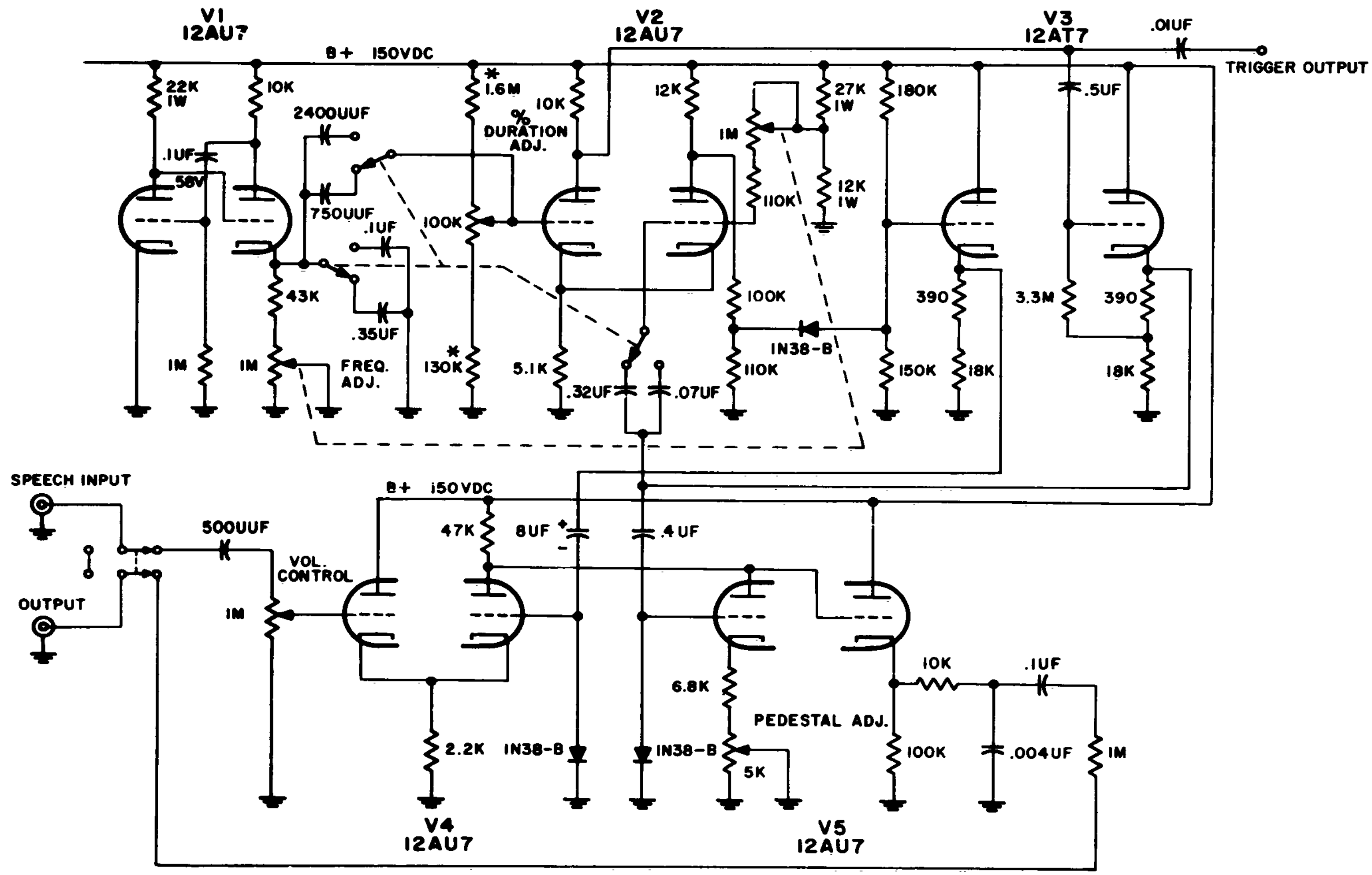
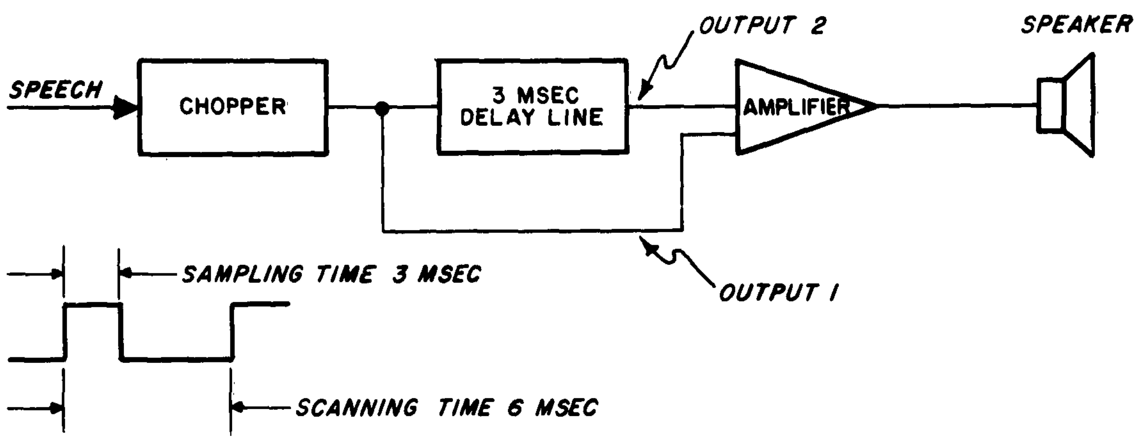


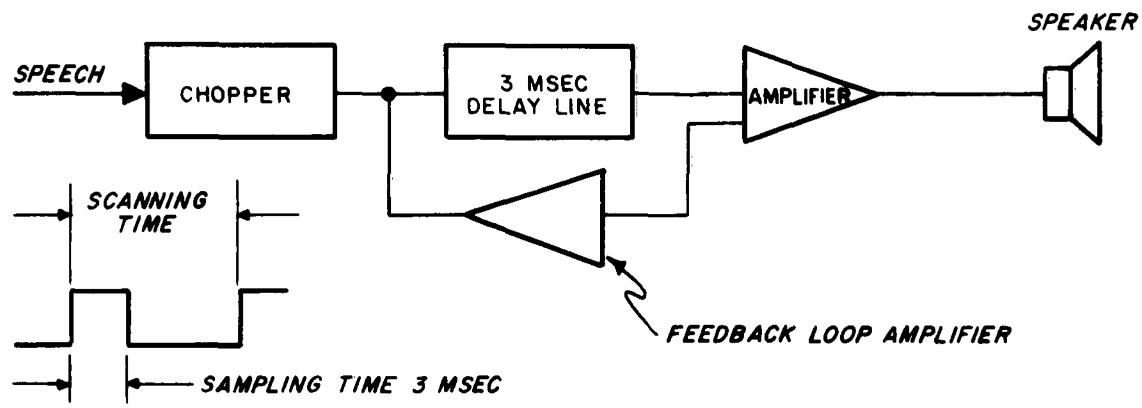
Figure 27. Schematic of Time Sampler or Speech Chopper



Data on Time-Sampled Speech (3/6 msec) Using Delay Line

Delay Line	Degree Of Reverberation	Articulation Score	Remarks
0	0	84	Interrupted speech
3	0	71	Three-millisec delay
3	0	75	line to reconstruct
3	0	74	chopped speech

Figure 28. Test Setup and Data on Time-Sampled Speech (3/6 msec) Using Delay Line Reverberator



Data On Time-Sampled Speech (3/15 msec) Using Open Loop Gain

Open Loop Gain AB	Average Articulation Score
0	56
0.2	44
0.5	59
0.78	36

Figure 29. Test Setup and Data On Time-Sampled Speech (3/15 msec) Using Feedback Reverberator

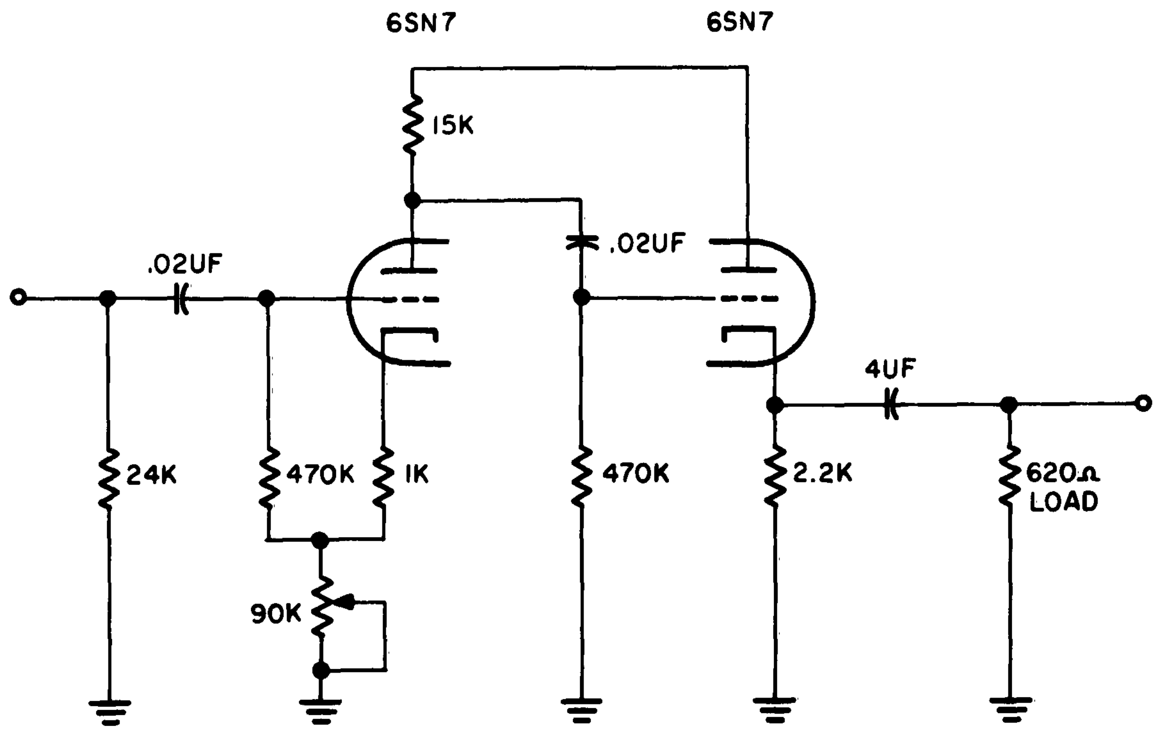


Figure 30. Schematic of Amplifier for Closed Loop Reverberator

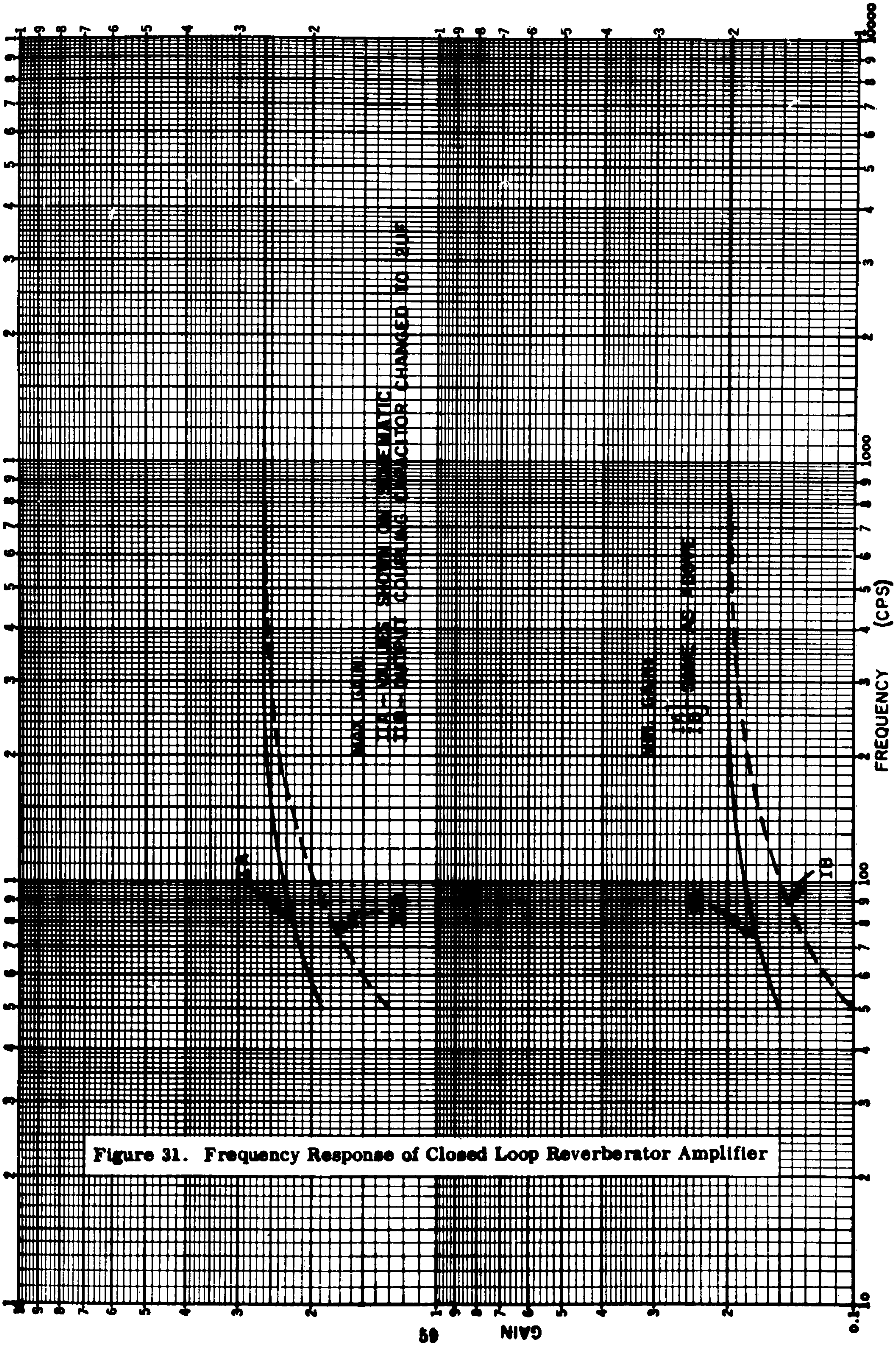


Figure 31. Frequency Response of Closed Loop Reverberator Amplifier

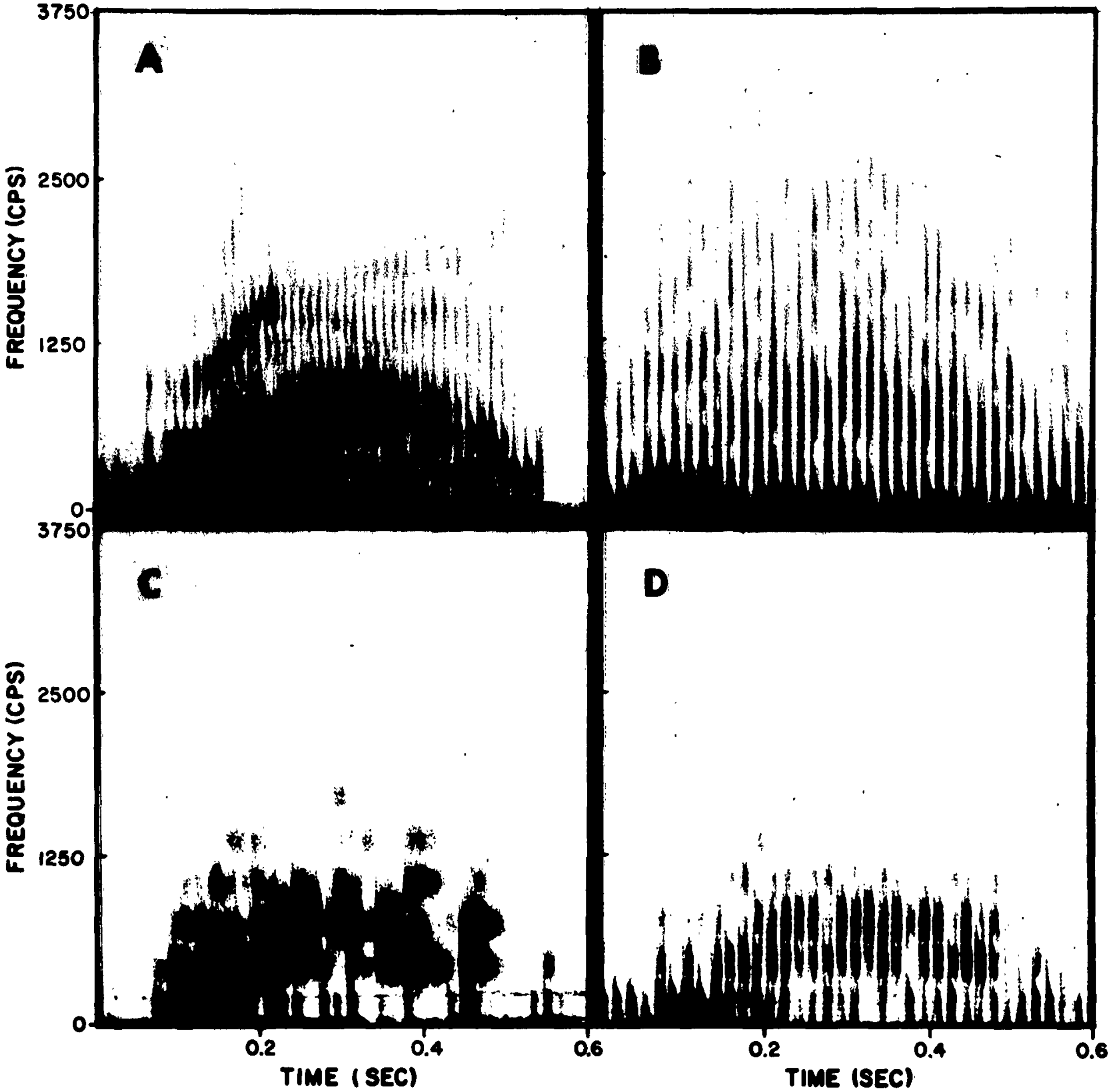


Figure 32 Speech Spectrograph "Rag" (a) Undistorted; (b) Chopped; (c) Chopped with Feedback A β=0.780; (d) Chopped with Feedback A β=0.200

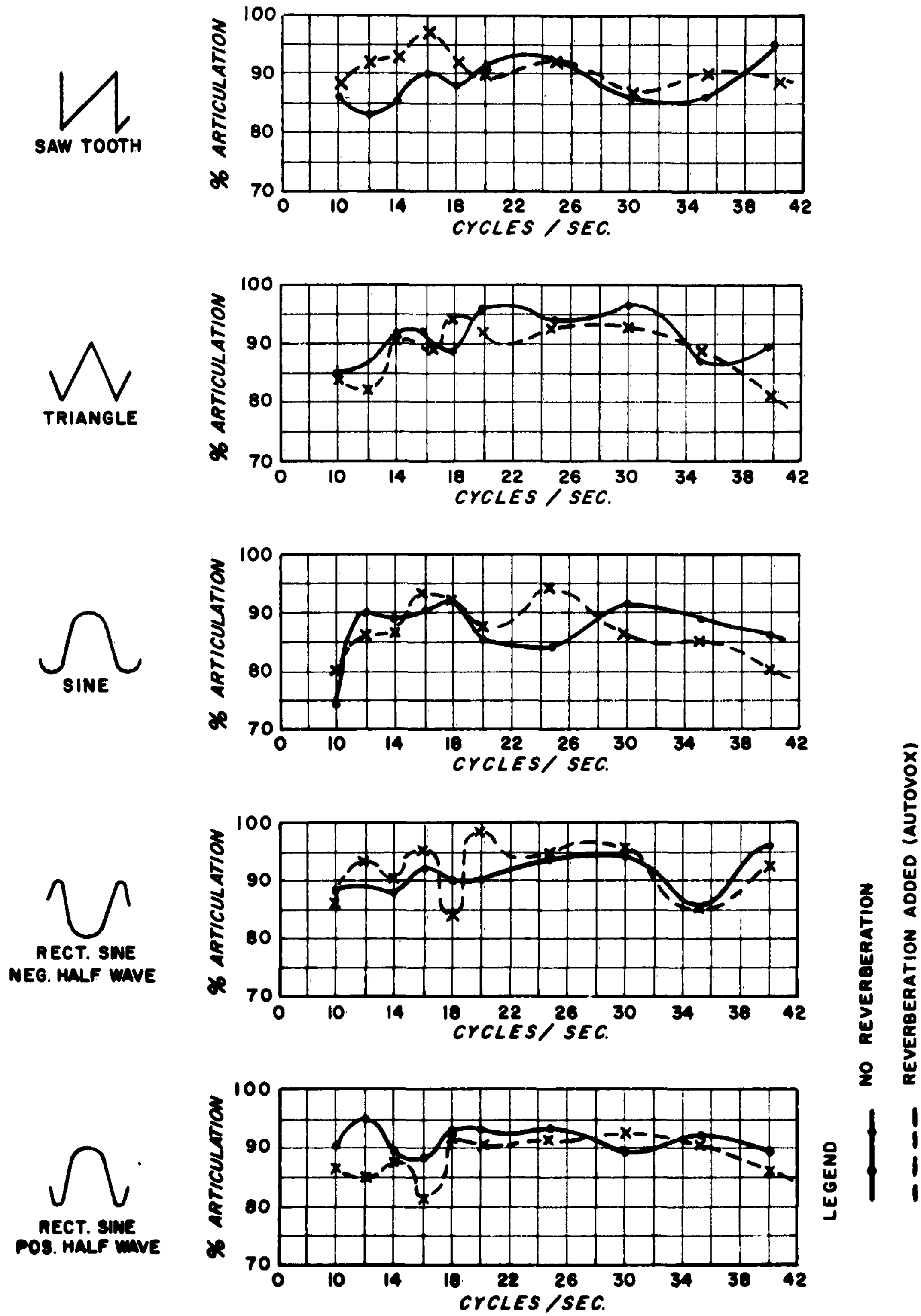


Figure 33. Average Articulation Scores for Time-Frequency Scanning, Compression Ratio 2:1

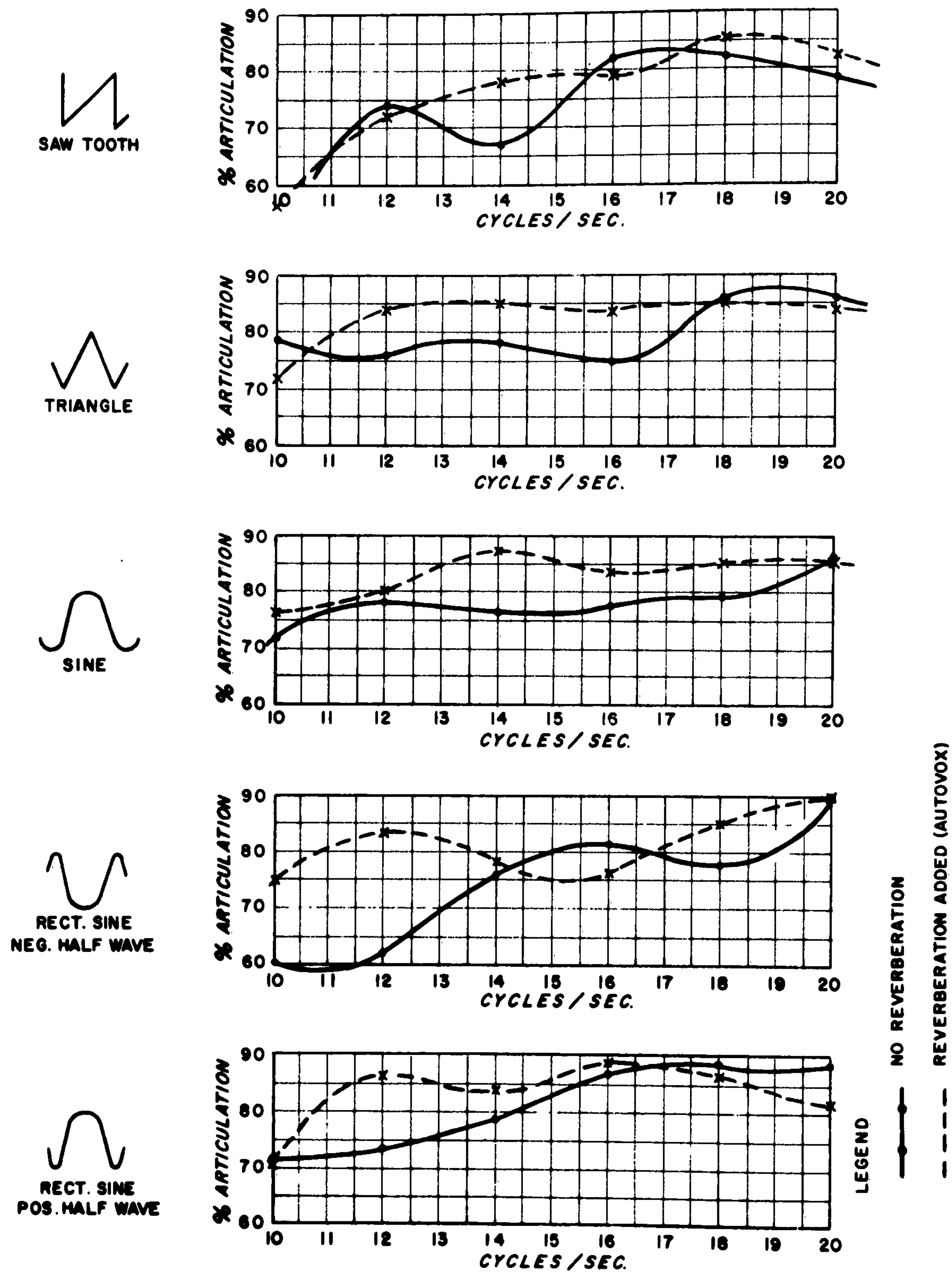


Figure 34. Average Articulation Scores for Time-Frequency Scanning, Compression Ratio 4:1

APPENDIX A

TECHNICAL PROPOSAL

Speech Bandwidth Compression
by Two Scanning Filters

NOTE: This proposal contains proprietary information which should not be divulged to other persons without the consent of PRD Electronics, Inc.

7 July 1961

P R D E L E C T R O N I C S, I N C.
A Subsidiary of Harris-Intertype Corporation
202 Tillary Street, Brooklyn 1, N. Y.

TABLE OF CONTENTS

MOTIVATION

This proposal describes a relatively simple method for compressing speech bandwidth by using two scanning filters. The ideas in this text have been generated from work performed on a Filter Scanning Speech Bandwidth Compression System sponsored by the Fort Monmouth Speech Processing Section on Contract No. DA-36-039-SC-85-140. The preliminary technique had been described to Messrs. M. Weinstock, J. DeClerk and F. Evans of Fort Monmouth, New Jersey during a visit to PRD Electronics on 25 May 1961.

In a subsequent visit to Bell Telephone Laboratories, Murray Hill, New Jersey, on 13 June 1961, Mr. J. DeClerk of Fort Monmouth and Mr. A. P. Albanese of PRD Electronics presented the proposed ideas to Drs. E. E. David,Jr., J. L. Flannagan, R. Miller, G. Raisbeck and M. R. Schroeder. This visit was to ascertain the originality of the system and to determine the validity of the system. The results were favorable because no system of its kind had been fabricated, and the ideas have promise for successful compression of speech bandwidth.

TECHNICAL DISCUSSION ON PROPOSED SYSTEM

The proposed system shown in Figure 1 is a simple speech bandwidth compression device capable of sampling, transmitting and restoring those parts of speech which are correlated to the articulatory function of speech. Since the word articulation can be attributed to the movements of the bars or formants, then the proposed transmitter extracts and transmits this information and the proposed receiver restores the spectrum into intelligible sounds. The basis for speech bandwidth compression can best be understood by observing a typical sound spectrogram for voiced sounds shown in Figure 2. Here, the transmitter narrow band pass filters are continuously scanning the

formant movements. The resulting signal to be transmitted is composed of pertinent spectrum samples corresponding to the information content of speech. On the average, the transmitted signal will contain both the first and second formant frequencies. However, when third formant energies are dominant, they will occasionally replace the lower intensity second formants. Although the ideal bandwidth for the first two formants are 7.1 and 6.7 cps, respectively[1], the effectiveness in tracking these formants will dictate the allowable bandwidth of the narrow band pass filters. It should be noted that the reception of the first two formants are sufficient to maintain high intelligibility.

The method proposed for sampling the formant movements is to generate oscillator control voltages from the speech spectrum. Figure 3a illustrates a short term power spectrum whose peak values correspond to the bars or formants shown in Figure 2. Integration of the speech signal will result in a slowly varying waveform which can be viewed in the frequency domain as low pass filtering shown in Figure 3b. Alternatively, differentiation is analogous to high pass filtering as seen in Figure 3c. If integrated and differentiated speech signals are infinitely clipped, then the resulting waveforms will be a binary array of random pulse widths and periods as shown in Figure 4. Due to the filtering aspects, it is obvious that the number of zero crossings for differentiation is greater than that of the integrated case. This is an important observation because the number of zero crossings is a measure of the average frequency of these waveforms. Therefore, the integrated and differentiated cases are respectively correlated to the first and second formant frequencies. Converting the average frequency of these waveforms into voltage amplitude via FM detection will result in amplitude variations corresponding to frequency shifts of the peak energies (see

the enclosed Appendix).

As seen in Figure 1, the output of the discriminators are used to control the frequency of the scanning oscillators. Although modulating the speech with the scanning oscillator yields two sideband frequencies with suppressed carrier, the narrow band pass filters transmit only the difference frequencies resulting in single sideband transmission. This condition can be depicted as though the filters were scanning the speech spectrum as shown in Figure 2. The transmission signal can be viewed as simply translating the speech spectrum up to the transmission pass band. The control voltages are also modulated or translated to the pass band of the facility. It is expected that the bandwidth of these control voltages will be small, something in the order of 10 to 20 c.p.s. each.

At the receiver, the compressed signal and control voltage information will be processed by the inverse action of the transmitter. Frequency selectivity by filtering is used to separate the four basic signal components as shown in Figure 1. The extracted control voltages are used in conjunction with the formant compressed signals for demodulation. This process can be visualized as retranslating the formant compressed signal frequencies to audible sounds. Since the audible sounds are the difference frequencies, it is questionable whether or not low pass filtering is required because the sum frequencies will lie outside the audible spectrum. The output of both demodulators are combined to restore the formant spectrum.

PRELIMINARY RESULTS

A crude test has been performed on differentiated and infinitely clipped speech. The purpose of this test was to compare the spectrum of undistorted speech to that of processed speech. The schematic diagram of Figure 5 is a differentia-or, amplifier, and limiter. A speech signal applied to the input

of this circuit can be viewed at the output as the random series of on-off pulses shown in Figure 4b. Using this circuit in conjunction with a Missilizer, results in the spectrogram shown in Figure 6b. Comparing the two spectrograms of Figure 6, it can be seen that the first formant for processed speech has been attenuated and the second formant emphasized. This agrees quite well with the anticipated system behavior described herein.

PROPOSED PROGRAM

A program of research and development is proposed leading to an experimental model incorporating the speech compression ideas herein disclosed. Pertinent questions to be answered by such a program are:

1. FM Detector Time Constant:

   Here it is required to choose a suitable time constant so that the amplitude variations of the control voltages are effectively averaging the formant movements. It can be seen that extreme time constant values will produce impulse and d.c. variations which are not applicable. Since there is an optimum detector time constant for the differentiated and integrated case, it will be our objective to select both time constants for optimum system performance.

2. Synchronization:

   Due to the continual frequency variation of the scanning oscillators, the transmitter and receiver must be correctly synchronized to ensure proper restoration of the speech spectrum. One difficulty usually encountered with systems similar to the proposed system is the non-linear phase characteristic of the narrow band pass filters. The presence of phase distortion will result in time delay variation as

a function of frequency within the filter pass band. At the receiver the reconstructed spectrum will be shifted in frequency by an amount proportional to the frequency location. The resulting word articulation scores will be adversely affected due to poor spectrum restoration. It will be our objective to rectify this matter by accounting for distortions present in the filter response by phase compensation and/or, if possible, obtaining filters with correct phase response.

3. Compression Ratio:

It is proposed that the bandwidth of the narrow band filters be approximately 200 c.p.s. each. The reason for this choice is to determine the effectiveness of this degree of bandwidth compression. The compressed signal pass band is expected to be 200 + 200 = 400 c.p.s., and the control voltages will have a combined bandwidth of 40 c.p.s. Therefore, for a total transmission bandwidth of 440 c.p.s., the expected degree of compression is almost $\frac{3000}{440} \approx 7$. The results obtained herein will determine whether or not further compression ratios are advisable.

It should be appreciated that the foregoing questions can only be answered by experimental tests not answered by a priori by analysis since the ear brain function is largely unknown and the properties of speech are unpredictable.

A program of about one year with two to three engineers and technical shop services is considered sufficient to arrive at an experimental model and to provide results of the expected degree of compression. Since

PRD Electronics has been actively engaged in speech communication systems; such as, speech bandwidth compression, voice privacy, and problems related to the transmission of binary speech codes, our facility in these areas have continually expanded. Many technical personnel have been trained in basic communication problems, so that their services can be applied to systems similar to the proposed device.

APPENDIX

It has been shown by Licklider[2] that integrated and differentiated infinitely clipped speech articulation scores can be related to zero crossing information. Observing a short term power spectrum for a typical vowel sound shown in Figure 3a, we can resolve the effects of differentiation and integration by viewing the result as low and high pass filtering as shown in Figure 3b and 3c respectively. Obviously formant separation can be obtained if sharp filters are used. The infinite clipping of both signals then resolve speech information into the location of the zero crossings. Since infinitely clipped speech signals are a series of square waves of uniform amplitude, the information must lie in the average frequency variation or zero crossings. If one considers an FM signal at the output of a limiter, the waveform also constitutes a series of square waves which have frequency variation. Upon injecting this signal into an ideal discriminator, the output voltage will be directly proportional to the average instantaneous frequency. Here, we can deduce that the average frequency of differentiated and infinitely clipped speech will be greater than that of integrated and infinitely clipped because of the filtering aspects. It is this condition that correlates the integrated and differentiated waveforms to the first and second formants respectively.

Consider the ideal case of an FM signal shown in Figure 7a, and assume the modulated signal is a sine wave. The mathematical description of this wave is well known as[3]

$$v(t) = A \cos (W_0 t + \ \sin W_m t) = A \cos \Theta (t) \qquad (1)$$

where

A is a constant equal to the amplitude of the waveform,

W_0 is the carrier frequency in radious/sec.

β is the modulation index, i.e. the ratio of frequency variation

to the modulation frequency, and

W_m is the modulation frequency.

The characteristics of an ideal discriminator will provide an output voltage proportional to $\frac{d\Theta(t)}{dt}$ namely,

$$V_{out}(t) = W_o + \Delta W \cos W_m t \text{ volts} \tag{2}$$

The FM detector output (omitting the dc term, W_o) will be a voltage signal proportional to the frequency of the carrier. If the signal is limited so that it appears as shown in Figure 7b, then an alternate solution for its behavior must be used. The technique introduced by Stompers(4) which makes use of the average zero crossings of the wave is applicable to both waves shown in Figures 7a and 7b and is also valid for waveforms which are perturbed by external sources, e.g. noise and speech variation. Stompers defines the instantaneous frequency as

$$W_i = \frac{\pi}{T} \times \text{no. of zero crossings in T sec.}$$

$$= \frac{2\pi}{T} \times \text{no. of positive zero crossings in T sec.}$$

Where T is chosen in the interval $\frac{1}{f_o} < T < \frac{1}{B}$

and B is the signal bandwidth.

The process of FM detection can be visualized by considering the counting function of Figure 8. For each successive zero crossing, the counting function increases one unit. The pattern displayed can be found by evaluating the zero crossings of equation (1) viz.

$$W_o t + \beta \sin W_m t = (2n-1) \frac{\pi}{2} \tag{3}$$

which can be solved graphically as shown in Figure 9. The time at which a zero will occur is determined by the intersection of $\beta \sin W_m t$ and the parametric $(2n - 1) \frac{\pi}{2} - w_o t$ equations. An alternate solution for the counting function could have been obtained from equation (3) if we plot the left side as a continuous function and then choose discrete unit values along the

ordinate. From these figures, the output voltage of the FM detector is $\frac{\Delta n}{\Delta t}$ yielding the required average frequency of the FM waveforms.

Extending this idea to differentiated and/or integrated and infinitely clipped speech, the discriminator process can be exploited by observing Figure 10. Assume a typical speech pattern shown in Figure 10a. It is desired to obtain a voltage amplitude measure of the average frequency of this waveform as a function of time. FM detection will provide output voltage proportional to the average frequency within a time interval T. For example, the process the infinitely clipped signal follows is differentiation, as shown in Figure 10b and detection shown in Figure 10c. The detector low pass filter output voltage will be proportional to the number of impulse functions within the time constant interval as can be seen in Figure 10d. Therefore, the output waveform is a slowly varying signal, correlated to the average frequency of the speech zero crossings. This signal will be applied to the scanning oscillator of the system to scan the corresponding speech formants as shown in Figure 1.

REFERENCES

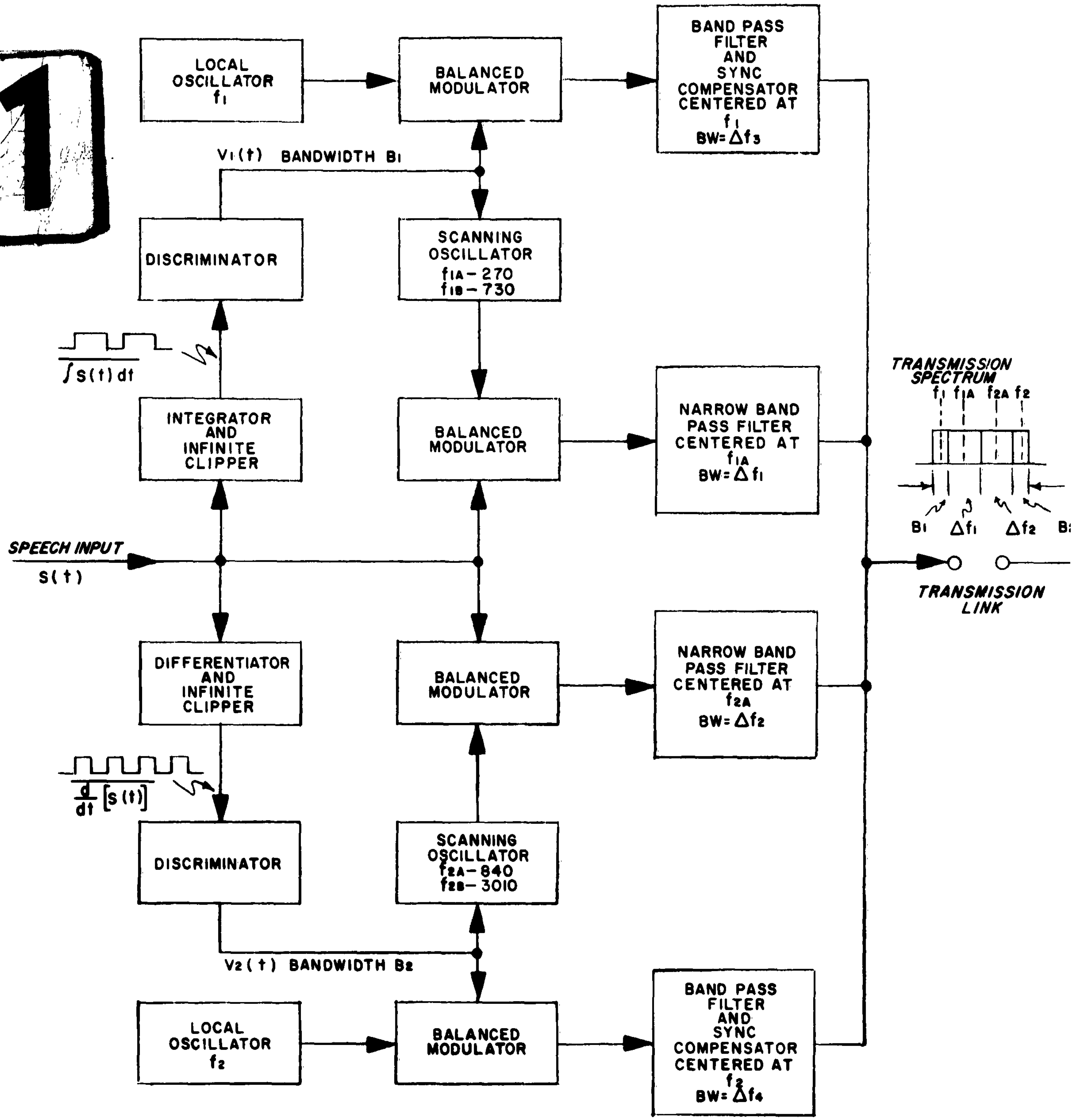
(1) Flannagan, J. L., "Bandwidth and Channel Capacity Necessary to Transmit the Formant Information of Speech", JASA, 1956, Vol. 28, No. 4, July 1956, pp. 592 - 596.

(2) Licklider, J. C. R. and Pollack, I.: "Effective on Differentation Integration and Infinite Clipping upon the Intelligibility of Speech", JASA, 1948, Vol. 20, p. 42.

(3) H.S. Black: "Modulation Theory" D. Van Nostrand Co., Inc., 1953.

(4) Stumpers, F. L. H. M.: "Theory of Frequency Modulation Noise", Proc. IRE, Sept. 1948, pp. 1081 - 1092.

1



LOCAL OSCILLATOR f1
BALANCED MODULATOR
BAND PASS FILTER AND SYNC COMPENSATOR CENTERED AT f1 BW= Δf3
V1(t) BANDWIDTH B1
DISCRIMINATOR
SCANNING OSCILLATOR f1A – 270 f1B – 730
∫s(t) dt
INTEGRATOR AND INFINITE CLIPPER
BALANCED MODULATOR
NARROW BAND PASS FILTER CENTERED AT f1A BW= Δf1
TRANSMISSION SPECTRUM
f1 f1A f2A f2
B1 Δf1 Δf2 B2
SPEECH INPUT
S(t)
TRANSMISSION LINK
DIFFERENTIATOR AND INFINITE CLIPPER
BALANCED MODULATOR
NARROW BAND PASS FILTER CENTERED AT f2A BW= Δf2
d/dt [s(t)]
DISCRIMINATOR
SCANNING OSCILLATOR f2A – 840 f2B – 3010
V2(t) BANDWIDTH B2
LOCAL OSCILLATOR f2
BALANCED MODULATOR
BAND PASS FILTER AND SYNC COMPENSATOR CENTERED AT f2 BW= Δf4

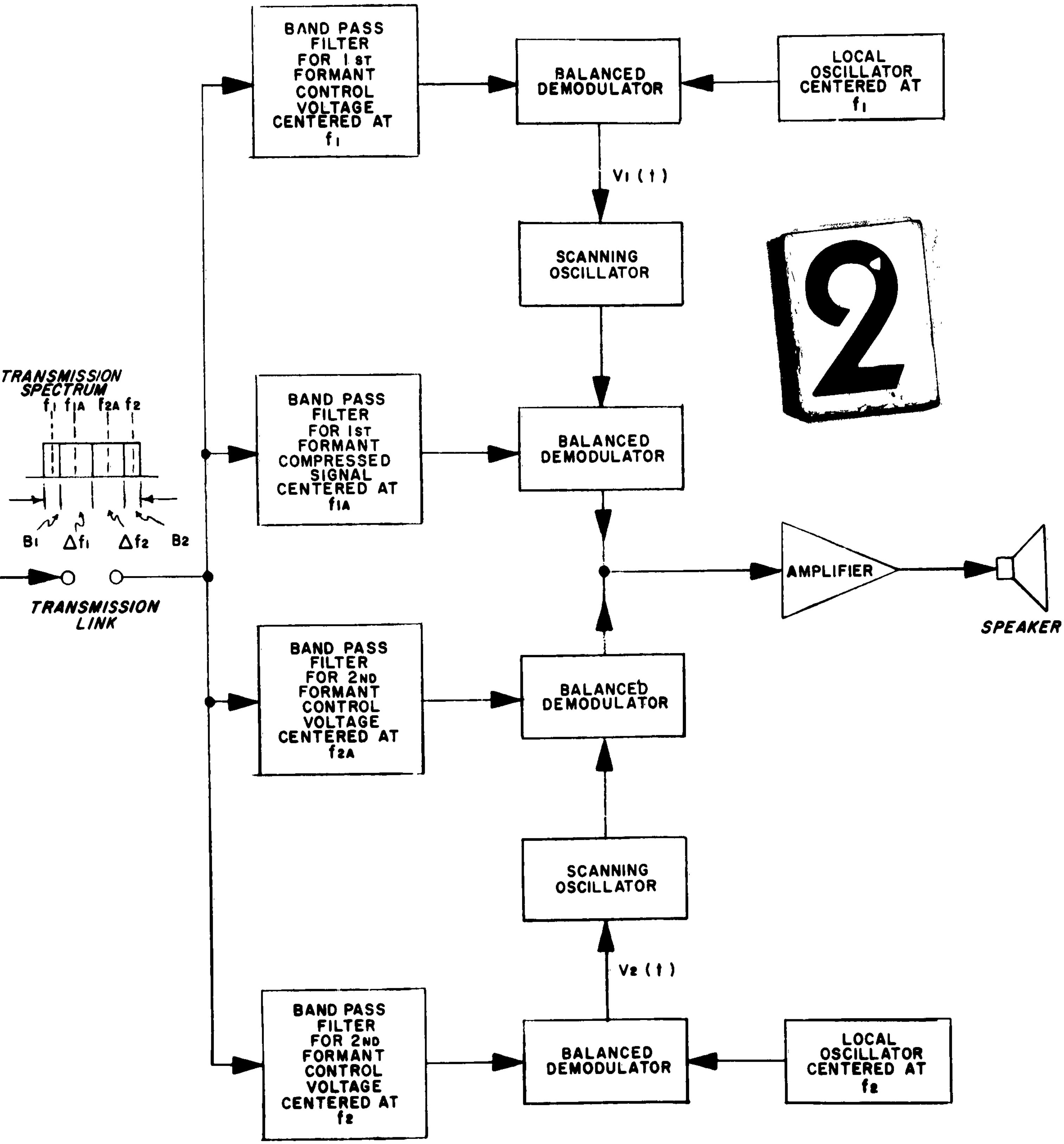
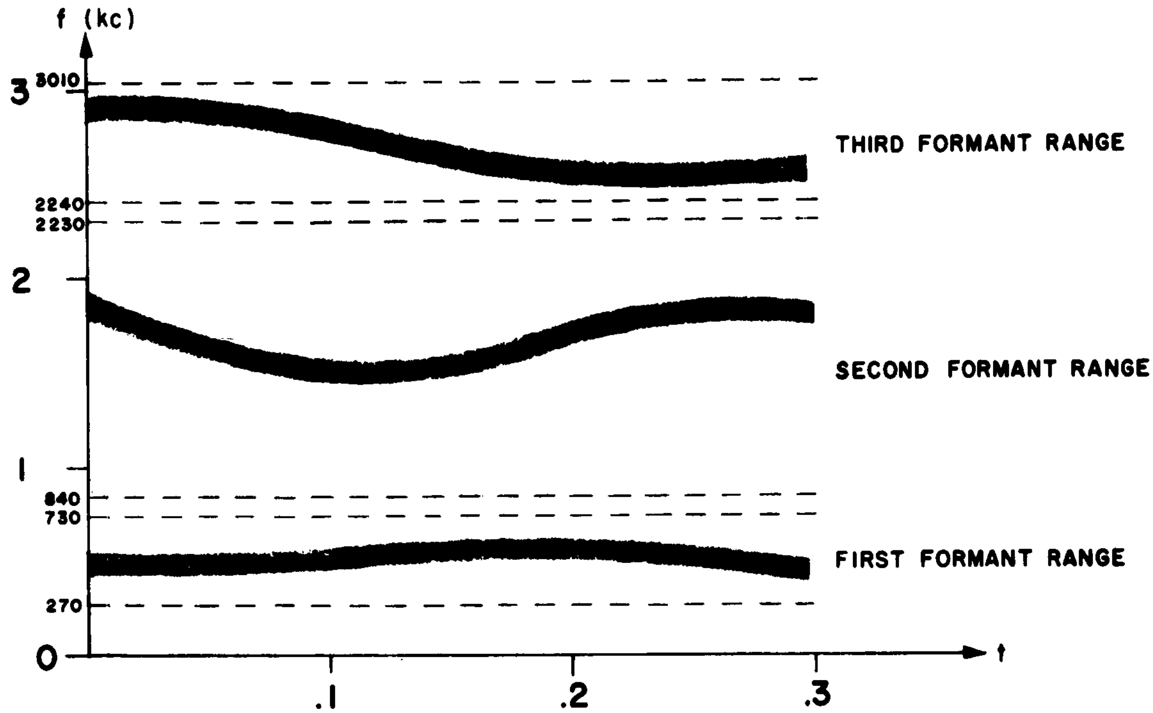
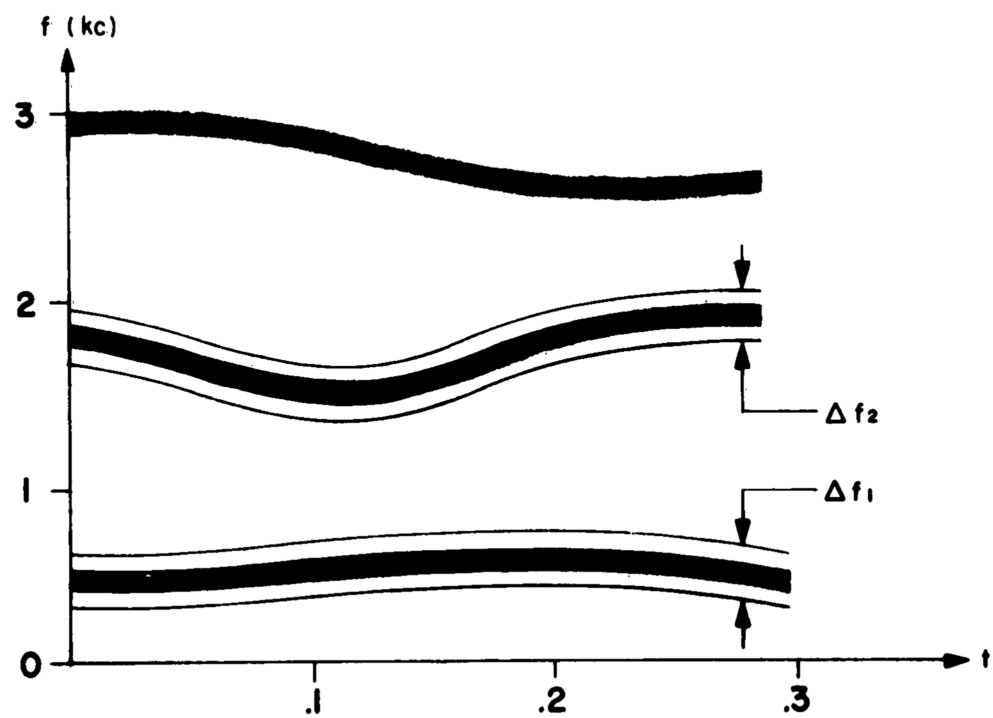


Figure 1 BLOCK DIAGRAM OF PROPOSED SPEECH BANDWIDTH COMPRESSION SYSTEM



(a) Sound Spectrogram of Voiced Speech Signal



(b) Narrow Band Pass Filter Scanning

Figure 2 SOUND SPECTROGRAM EXTRACTING FORMANT MOVEMENTS FOR TRANSMISSION SPEECH SIGNAL

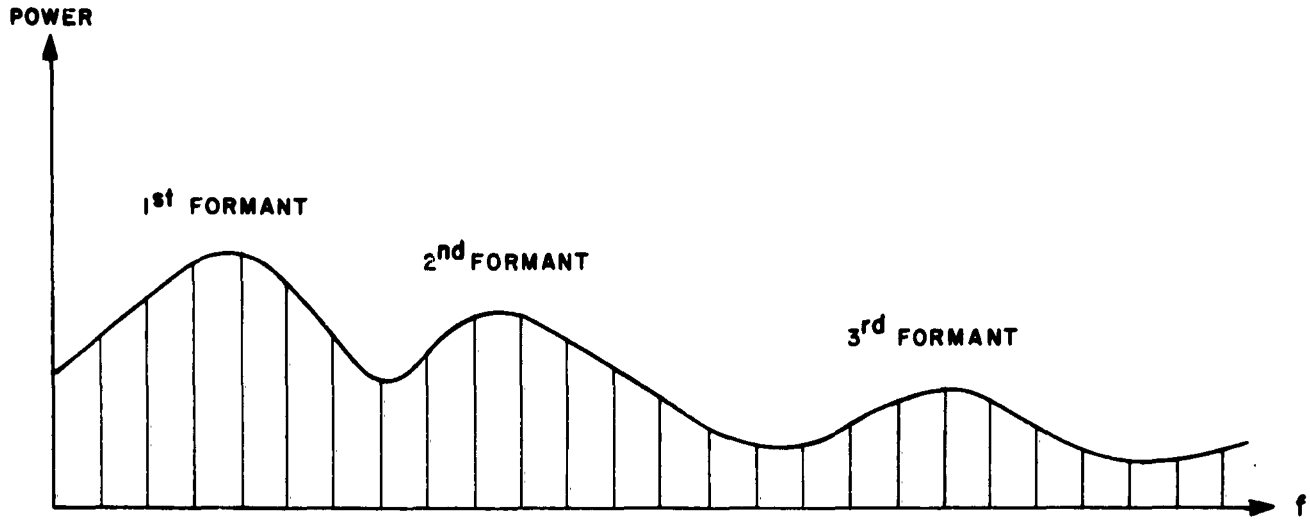


Figure 3a SPECTRAL DISTRIBUTION OF SPEECH ENERGY

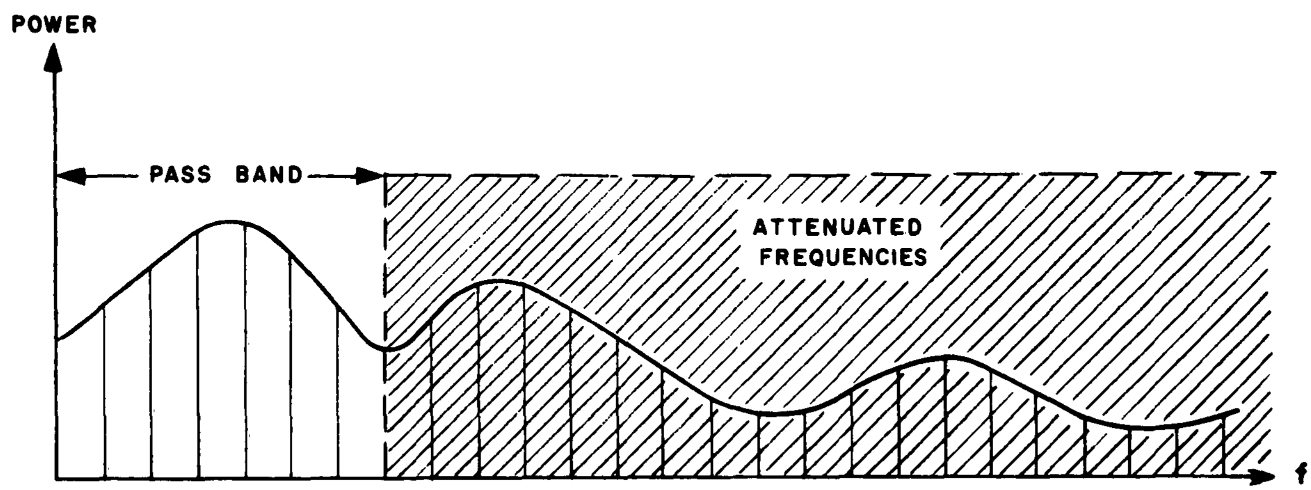


Figure 3b EFFECT ON SPECTRUM WHEN INTEGRATING SPEECH SIGNAL

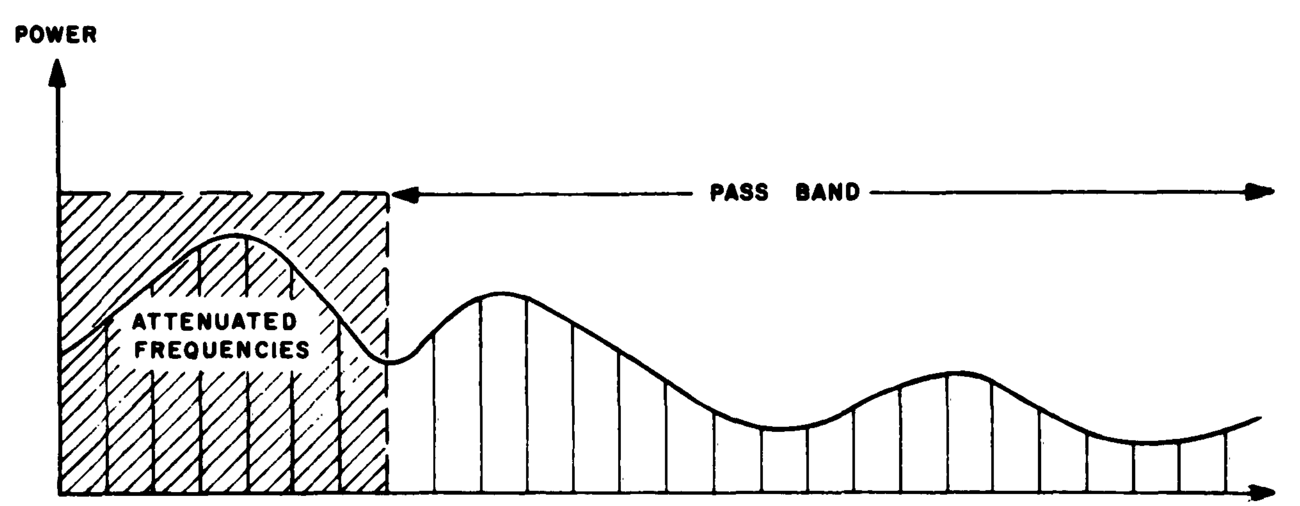


Figure 3c EFFECT ON SPECTRUM WHEN DIFFERENTIATING SPEECH SIGNAL

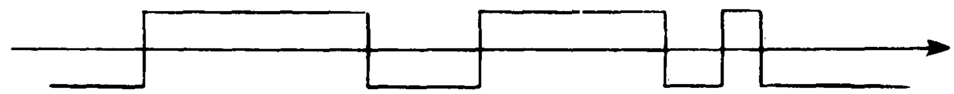


Figure 4A
INTEGRATED AND INFINITELY CLIPPED SPEECH

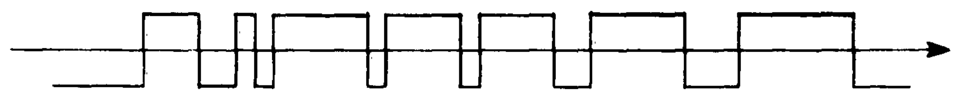


Figure 4B
DIFFERENTIATED AND INFINITELY CLIPPED SPEECH

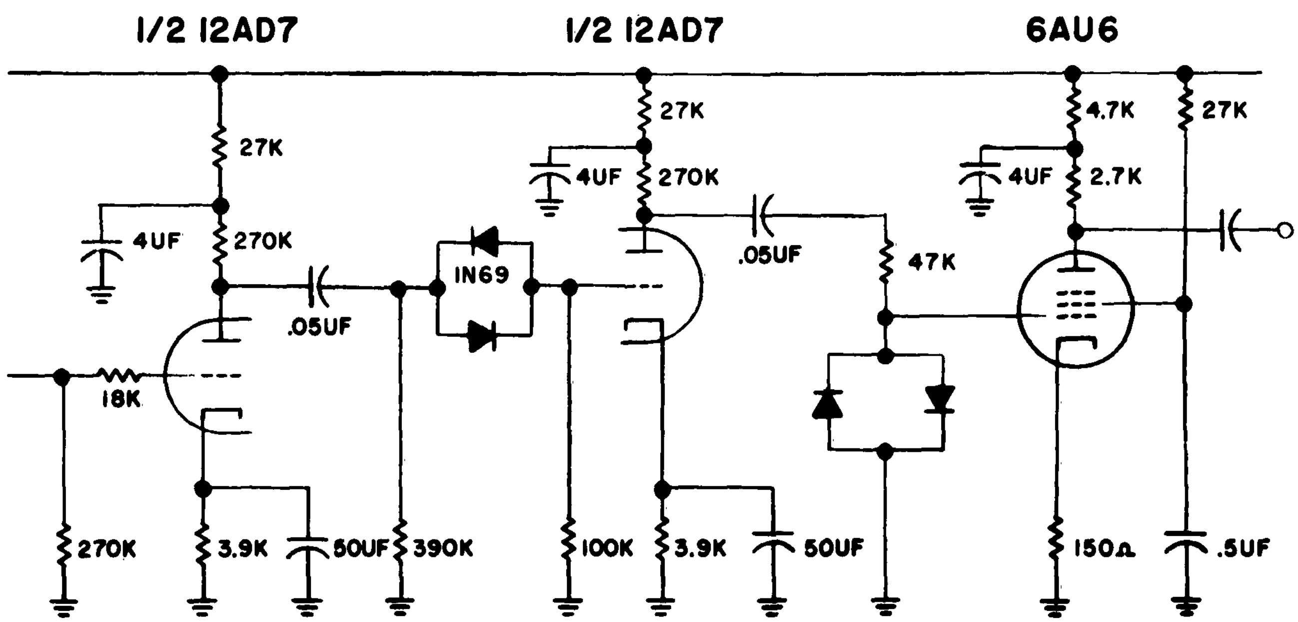


Figure 5
DIFFERENTIATION AND CLIPPING SIGNAL AMPLIFIER

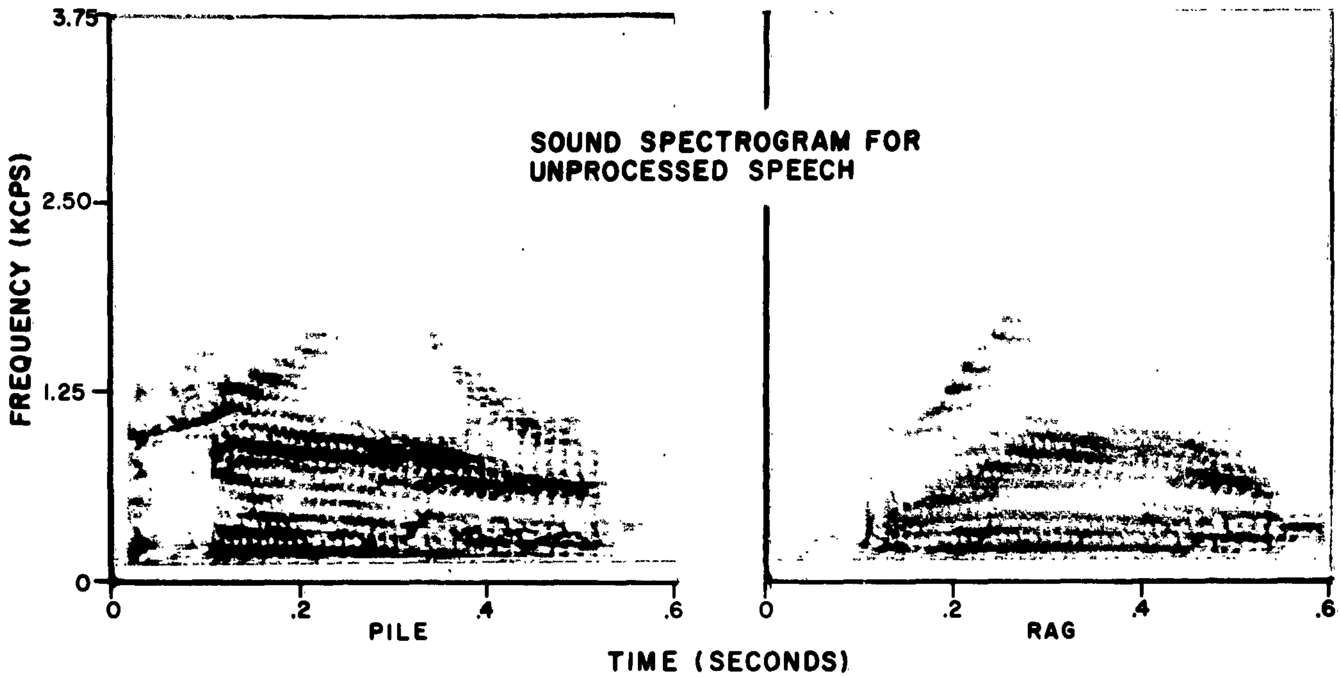


Figure 6A

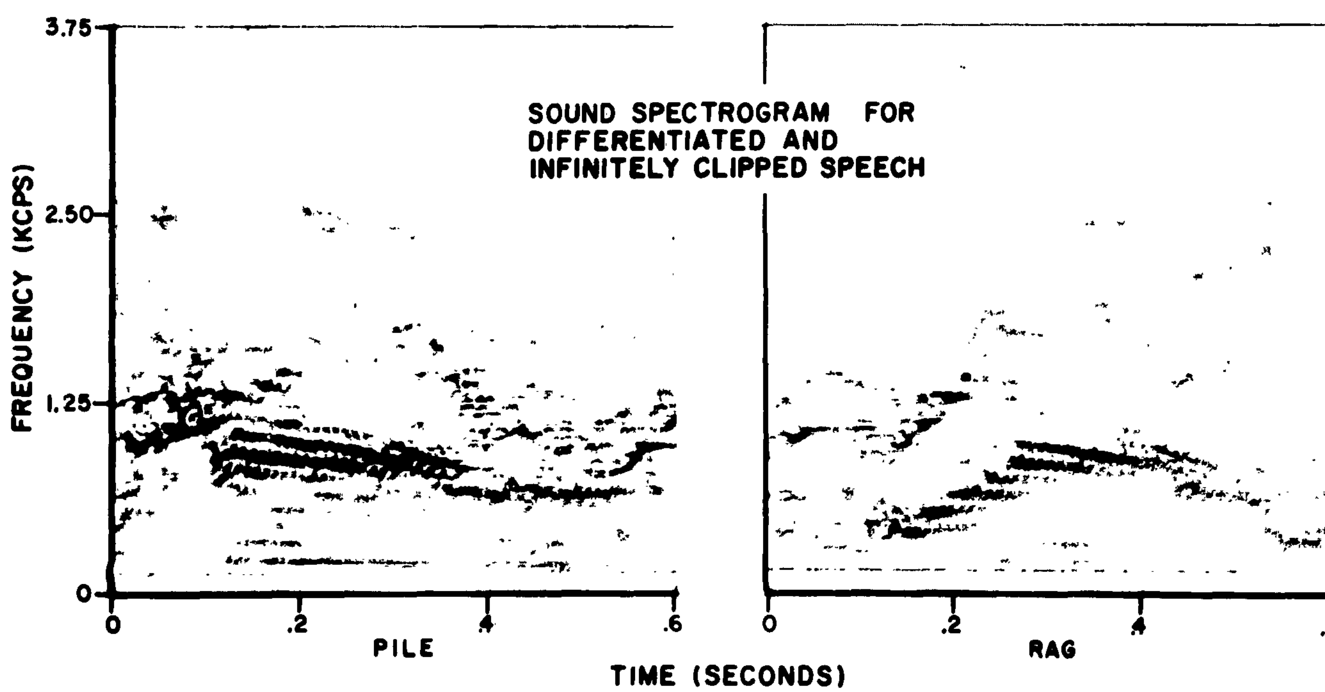


Figure 6B

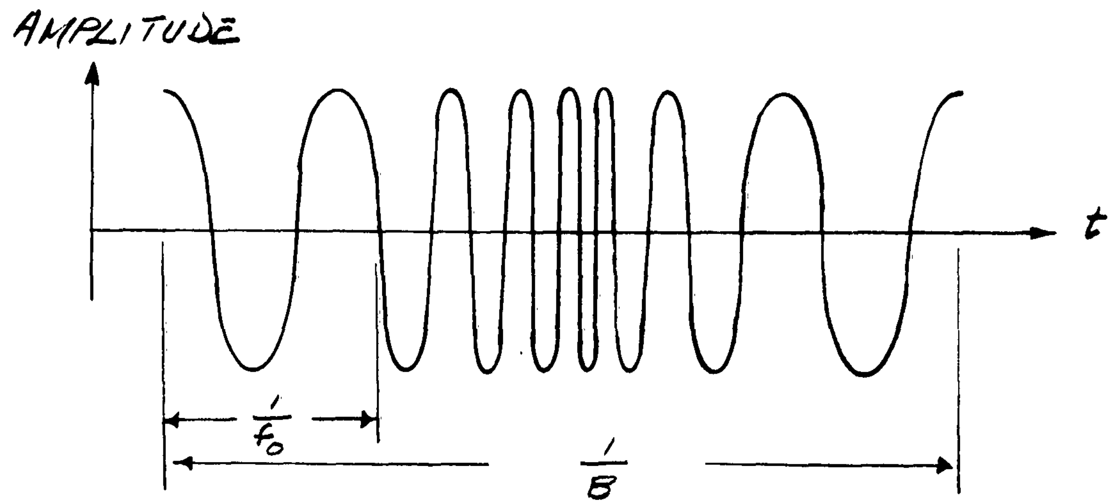


Figure 7a FREQUENCY SINE-WAVE MODULATED WAVEFORM

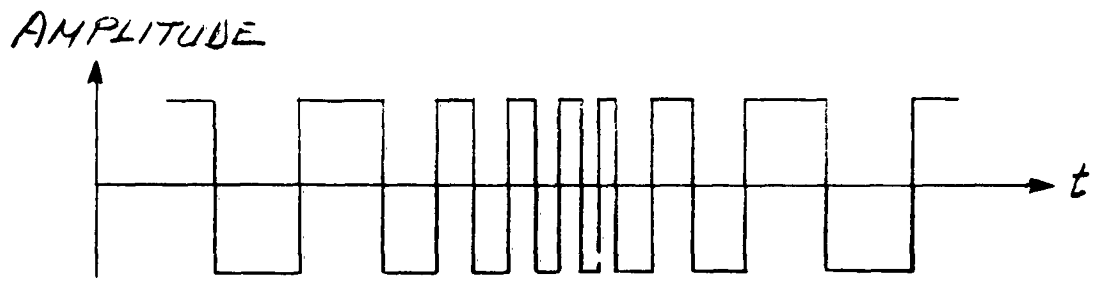


Figure 7b INFINITELY CLIPPED FM WAVEFORM

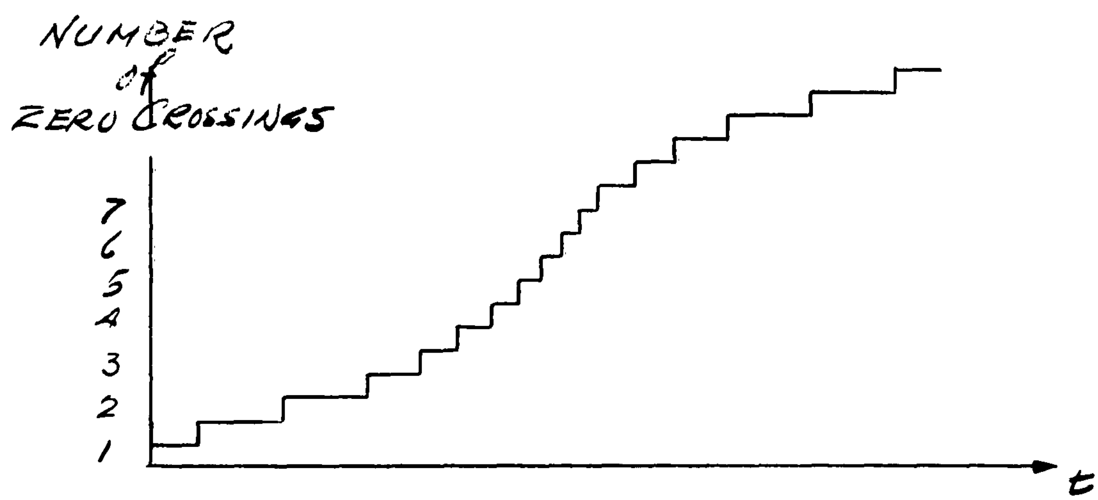


Figure 8 COUNTING FUNCTION vs LINE

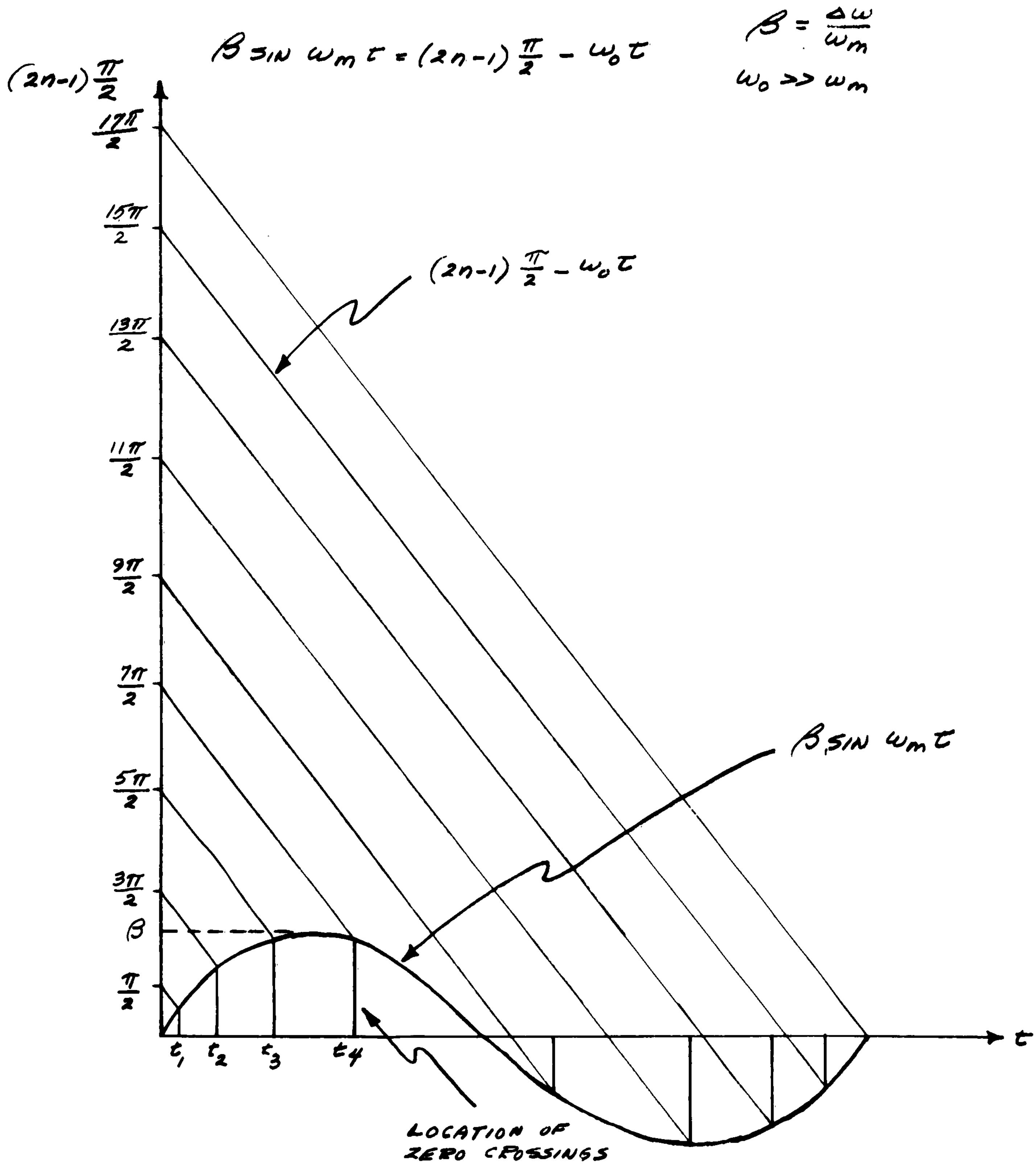
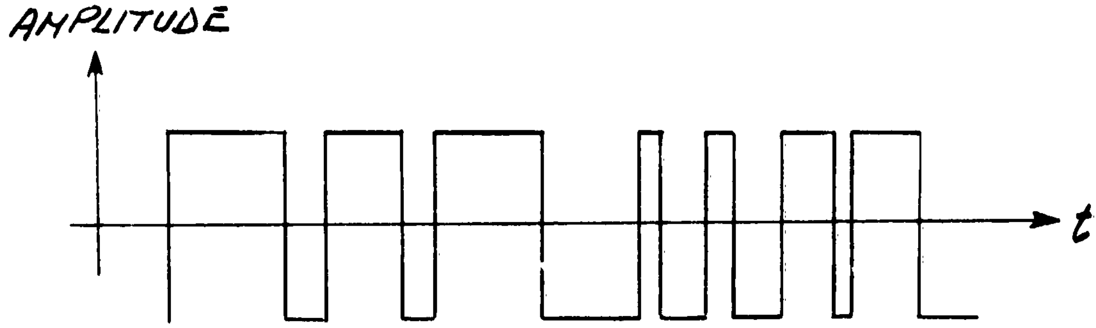
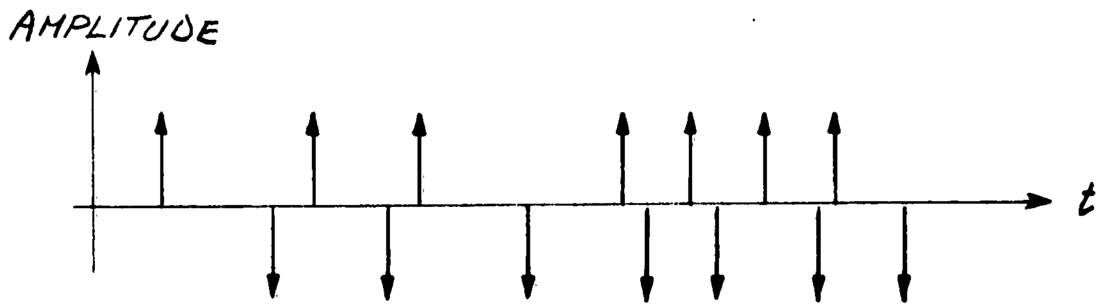


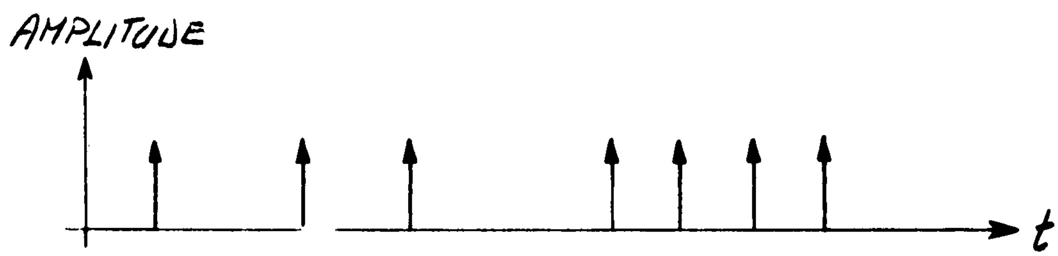
Figure 9 GRAPHICAL SOLUTION FOR DETERMINATION OF ZERO CROSSINGS



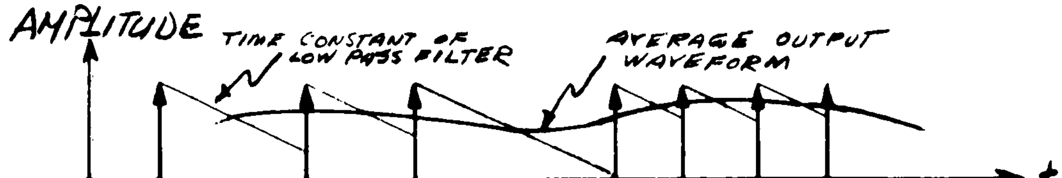
(a) INFINITELY CLIPPED SPEECH



(b) DIFFERENTIATED INFINITELY CLIPPED SPEECH



(c) DETECTED DIFFERENTIATED INFINITELY CLIPPED SPEECH



(d) DETECTOR OUTPUT

Figure 10 WAVEFORMS OF PROCESSED SPEECH

AD- Div. 5/5, 25/1

PRD Electronics, Inc., Sub. of Harris-Intertype Corporation, Brooklyn 1, N. Y.
SPEECH BANDWIDTH COMPRESSION SYSTEM, by A. P. Albanese, Final Report, for period of July 1, 1960 to September 30, 1961. 82 p. incl. illus., tables, 8 refs.
(Final Report)
(Contract DA-36-039-sc-85-140) Unclassified Report

This report describes the analysis, experimentation and fabricational activity directed toward a time-frequency scanning speech compression system. A statistical analysis was made to determine the scanning waveforms based on articulation scores for selected zero crossings of infinitely clipped speech. In addition, an empirical approach using sawtooth, triangular, sine-wave and rectified sine-wave scans was employed to determine the sampling scan and rates which would yield maximum articulation. Experimentation shows that using an Autovox for reverberation improved articulation. This project initiated the concept of a
(over)

UNCLASSIFIED

1. Speech analysis
2. Speech intelligibility
3. Speech transmission-effectiveness

I Army Signal Research and Development Laboratory, Fort Monmouth, N. J.
II Contract DA-36-039-sc-85-140

Armed Services Technical Information Agency
UNCLASSIFIED

AD- Div. 5/5, 25/1

PRD Electronics, Inc., Sub. of Harris-Intertype Corporation, Brooklyn 1, N. Y.
SPEECH BANDWIDTH COMPRESSION SYSTEM, by A. P. Albanese, Final Report, for period of July 1, 1960 to September 30, 1961. 82 p. incl. illus., tables, 8 refs.
(Final Report)
(Contract DA-36-039-sc-85-140) Unclassified Report

This report describes the analysis, experimentation and fabricational activity directed toward a time-frequency scanning speech compression system. A statistical analysis was made to determine the scanning waveforms based on articulation scores for selected zero crossings of infinitely clipped speech. In addition, an empirical approach using sawtooth, triangular, sine-wave and rectified sine-wave scans was employed to determine the sampling scan and rates which would yield maximum articulation. Experimentation shows that using an Autovox for reverberation improved articulation. This project initiated the concept of a
(over)

UNCLASSIFIED

1. Speech analysis
2. Speech intelligibility
3. Speech transmission-effectiveness

I Army Signal Research and Development Laboratory, Fort Monmouth, N. J.
II Contract DA-36-039-sc-85-140

Armed Services Technical Information Agency
UNCLASSIFIED

AD- Div. 5/5, 25/1

PRD Electronics, Inc., Sub. of Harris-Intertype Corporation, Brooklyn 1, N. Y.
SPEECH BANDWIDTH COMPRESSION SYSTEM, by A. P. Albanese, Final Report, for period of July 1, 1960 to September 30, 1961. 82 p. incl. illus., tables, 8 refs.
(Final Report)
(Contract DA-36-039-sc-85-140) Unclassified Report

This report describes the analysis, experimentation and fabricational activity directed toward a time-frequency scanning speech compression system. A statistical analysis was made to determine the scanning waveforms based on articulation scores for selected zero crossings of infinitely clipped speech. In addition, an empirical approach using sawtooth, triangular, sine-wave and rectified sine-wave scans was employed to determine the sampling scan and rates which would yield maximum articulation. Experimentation shows that using an Autovox for reverberation improved articulation. This project initiated the concept of a
(over)

UNCLASSIFIED

1. Speech analysis
2. Speech intelligibility
3. Speech transmission-effectiveness

I Army Signal Research and Development Laboratory, Fort Monmouth, N. J.
II Contract DA-36-039-sc-85-140

Armed Services Technical Information Agency
UNCLASSIFIED

AD- Div. 5/5, 25/1

PRD Electronics, Inc., Sub. of Harris-Intertype Corporation, Brooklyn 1, N. Y.
SPEECH BANDWIDTH COMPRESSION SYSTEM, by A. P. Albanese, Final Report, for period of July 1, 1960 to September 30, 1961. 82 p. incl. illus., tables, 8 refs.
(Final Report)
(Contract DA-36-039-sc-85-140) Unclassified Report

This report describes the analysis, experimentation and fabricational activity directed toward a time-frequency scanning speech compression system. A statistical analysis was made to determine the scanning waveforms based on articulation scores for selected zero crossings of infinitely clipped speech. In addition, an empirical approach using sawtooth, triangular, sine-wave and rectified sine-wave scans was employed to determine the sampling scan and rates which would yield maximum articulation. Experimentation shows that using an Autovox for reverberation improved articulation. This project initiated the concept of a
(over)

UNCLASSIFIED

1. Speech analysis
2. Speech intelligibility
3. Speech intelligibility-effectiveness

I Army Signal Research and Development Laboratory, Fort Monmouth, N. J.
II Contract DA-36-039-sc-85-140

Armed Services Technical Information Agency
UNCLASSIFIED

AD-

speech compression system using two filter scans to intercept formant changes and yield maximum articulation.

UNCLASSIFIED

UNITERMS

Speech
Compression
Scanning
Audio
Reverberation
Modulation
Transmission
Delay
Statistics, Speech
Zero Crossings
Infinitely Clipped
Speech

UNCLASSIFIED

AD-

speech compression system using two filter scans to intercept formant changes and yield maximum articulation.

UNCLASSIFIED

UNITERMS

Speech
Compression
Scanning
Audio
Reverberation
Modulation
Transmission
Delay
Statistics, Speech
Zero Crossings
Infinitely Clipped
Speech

UNCLASSIFIED

AD-

speech compression system using two filter scans to intercept formant changes and yield maximum articulation.

UNCLASSIFIED

UNITERIMS

Speech
Compression
Scanning
Audio
Reverberation
Modulation
Transmission
Delay
Statistics, Speech
Zero Crossings
Infinitely Clipped
Speech

UNCLASSIFIED

AD-

speech compression system using two filter scans to intercept formant changes and yield maximum articulation.

UNCLASSIFIED

UNITERMS

Speech
Compression
Scanning
Audio
Reverberation
Modulatior
Transmission
Delay
Statistics, Speech
Zero Crossings
Infinitely Clipped
Speech

UNCLASSIFIED

CONTRACT NO. DA36-039-SC-85140

DISTRIBUTION LIST

FINAL REPORT

Addressee	No. of Copies
OASD (R & E) Rm 3E1965 The Pentagon Washington 25, D. C. Attn: Technical Library	1
Chief Signal Officer Department of the Army Washington 25, D. C. Attn: SIGRD-6-e	1
Commander Air Force Cambridge Research Center L. G. Hanscom Field Bedford, Massachusetts Attn: CROTR	1
Commanding General U. S. Army Electronic Proving Ground Fort Huachuca, Arizona Attn: Technical Library	1
U. S. Continental Army Command Liaison Office U. S. Army Signal Research & Development Laboratory Fort Monmouth, N. J. Attn: SIGRA/SL-LNF	
Commanding Officer U. S. Army Signal Research & Development Laboratory Fort Monmouth, N. J. Attn: SIGRA/SL/NDV	
Commanding Officer U. S. Army Signal Equipment Support Agency Fort Monmouth, N. J. Attn: SIGFM/ES-ADJ	

Addressee	No. of Copies
Chief of Research and Development OCS, Department of the Army Washington 25, D. C.	1
Commander Wright Air Development Center Wright-Patterson Air Force Base Ohio Attn: WCOSI-3	2
Commander Rome Air Development Center Air Research and Development Command Griffiss Air Force Base, N. Y. Attn: RCSSLD	1
Commander Armed Services Technical Information Agency Arlington Hall Station Arlington 12, Virginia Attn: TIPDR	10
Commanding Officer U. S. Army Signal Research & Development Laboratory Fort Monmouth, N. J. Attn: SIGRA/SL-ADT	
Commanding Officer U. S. Army Signal Research & Development Laboratory Fort Monmouth, N. J. Attn: SIGRA/SL-TN (For retransmittal to accredited British and Canadian Government Representatives and to Department of Commerce.)	

Addressee	No. of Copies
Corps of Engineers Liaison Office U. S. Army Signal Research & Development Laboratory Fort Monmouth, N. J. Attn: SIGRA/SL-LNE	
Marine Corps Liaison Officer U. S. Army Signal Research & Development Laboratory Fort Monmouth, N J. Attn: SIGRA/SL-LNR	1
Commanding Officer U. S. Army Signal Research & Development Laboratory Fort Monmouth, N. J. Attn: Logistics Division (Marked for F. Evans, SIGRA/SL-NDV)	2
Deputy President U. S. Army Security Agency Board Arlington Hall Station Arlington 12, Virginia	1
Commanding Officer U. S. Army Signal Research & Development Laboratory Fort Monmouth, N. J. Attn: SIGRA/SL-DR, SIGRA/SL-DE	1
Chief, U. S. Army Security Agency Arlington Hall Station Arlington 12, Virginia	2

This contract is supervised by Voice Security Branch, Communications Security Division, Communications Department, U. S. Army Signal Research & Development Laboratory, Fort Monmouth, New Jersey. For further information contact Mr. F. Evans, Project Engineer. Telephone: 5351091.